
Um sistema de reconhecimento de objetos
incorporado a um robô humanoide com aplicação
na educação

Adam Henrique Moreira Pinto

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito: 03/12/2014

Assinatura:

Um sistema de reconhecimento de objetos incorporado a um robô humanoide com aplicação na educação

Adam Henrique Moreira Pinto

Orientadora: Profa. Dra. Roseli Aparecida Francelin Romero

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências - Ciências de Computação e Matemática Computacional. *EXEMPLAR DE DEFESA*

USP – São Carlos
Dezembro de 2014

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

P645s Pinto, Adam Henrique Moreira
Um sistema de reconhecimento de objetos
incorporado a um robô humanoide com aplicação na
educação / Adam Henrique Moreira Pinto; orientadora
Roseli Aparecida Francelin Romero. -- São Carlos,
2015.
78 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2015.

1. Robótica educacional. 2. Visão Computacional.
3. Atenção Visual. I. Romero, Roseli Aparecida
Francelin, orient. II. Título.

“A menos que modifiquemos a nossa maneira de pensar, não seremos capazes de resolver os problemas causados pela forma como nos acostumamos a ver o mundo”
(Albert Einstein)

Agradecimentos

É chegado o final de mais um ciclo, difícil como deveria ser, mas com certeza mais um momento para fortalecer amizades e entender de uma vez por todas que não estamos sozinhos neste mundo. Este espaço é pra agradecer a todos que me ajudaram nesta jornada. Não posso me dar ao luxo de esquecer um nome, então falarei dos grupos em geral. É sempre uma alegria para mim saber que muitas pessoas passaram pela minha vida, e quero lembrar de todas, mas precisaria de muito espaço para nomeá-las.

Em primeiro e acima de tudo, agradeço a Deus, pois toda manhã antes de começar meu dia é Nele que busco força, é Nele que encontro a paz.

À minha família, meus pais Renata e Josué e aos meus irmãos Josué Junior e Renan que me deram um lar, que me deram apoio, que me deram forças e que foram compreensíveis nos momentos mais complicados, e que estavam ao meu lado nas alegrias e nas trizezas. Um lar cheio de amor não vai nunca se desfazer.

À minha orientadora Roseli, por ter acreditado no meu trabalho, por ter me acompanhado e dado toda a base para que eu conseguisse chegar até aqui.

Ao Alcides, grande companheiro neste trabalho, que aguentou até reuniões as 22 horas de uma sexta-feira para me ajudar a concluir este trabalho.

À todos os professores da Universidade de São Paulo, que sempre me ajudaram com o que foi necessário, e que durante as disciplinas me ajudaram a entender esse mundo ao qual eu não acreditava que estava preparado. Em especial, gostaria de agradecer à professora Renata Meneguetti e ao Lucas Orlandi, pelo apoio na elaboração dos exercícios de matemática e na execução dos experimentos.

Aos amigos do Warthog Robotics, grupo de extensão que foi minha casa quando cheguei em São Carlos. Tive oportunidade, ao lado dessas pessoas, de realizar o sonho de competir na Robocup, e de ter ótimos resultados. Um abraço especial a todos que passaram por lá, que foram comer um lanche no Turbo e que trabalharam muito para sermos a grande equipe que somos.

Aos amigos do LAR/BioCom pela convivência diária. Todas as pausas para os cafés, todas as brincadeiras no email, a descontração anda de mãos dadas com a ciência neste laboratório, já que sempre pude também contar com a experiência e a inteligência deste pessoal para qualquer problema que surgisse.

À todos os demais amigos dos laboratórios da USP que conheci durante as disciplinas, festas, café, através de outros amigos, entre outros. Me ajudaram a me adaptar em São Carlos, comemoraram comigo as vitórias e me apoiaram quando parecia que as coisas não dariam certo. Durante este tempo de Mestrado fiz muitos amigos, devido a competições, conferências, minicursos ministrados. Quero agradecer a todos pela paciência e pelo convívio.

Aos amigos dos grupos de Teatro, os amigos do Futsal tanto aqui na USP quanto na UFSCAR. Quando a coisa realmente apertava o meu norte era saber que logo mais minha cabeça iria se preocupar com outra coisa, e eram vocês que estavam lá para me ajudar a desestressar. Isso foi muito importante para mim.

E claro, nunca vou esquecer de toda a base que deixei ao longo dos anos. À todos os familiares e amigos que fiz durante a vida e a graduação, meu eterno muito obrigado. Essas pessoas estavam sempre por perto graças à tecnologia, e sempre serem para mim um porto seguro, aonde sempre eu posso confiar.

Agradeço à FAPESP pela ajuda com o robô NAO, objeto importante nesta pesquisa

Por fim, agradeço à CAPES pelo apoio financeiro.

Resumo

Cada vez mais observa-se a inserção de novas tecnologias em salas de aulas. Com o auxílio de políticas públicas, computadores ligados a internet tem estado presentes nas classes nos mais longínquos lugares do nosso país. Isto tem proporcionado que o conhecimento chegue de forma mais ampla e irrestrita a todas as crianças em fase de desenvolvimento. Na última década, além de microcomputadores, tem-se observado a presença, em salas de aulas, de Ipads, celulares, cujos proprietários são os próprios alunos e até mesmo lousa eletrônica em escolas com poder aquisitivo maior. Aliado a isto, nota-se também a inserção de kits robóticos que tem motivado muito os alunos no aprendizado de raciocínio lógico e de programação, pois, eles experimentam o conceito: "aprender por meio do fazer". O uso de todas estas tecnologias tem como objetivo principal cativar a atenção dos alunos, incentivar a pesquisa e o aprendizado interativo, uma vez que o ensino antes expositivo dá lugar ao ensino interativo, isto é, que conta com a participação mais ativa do estudante. Nesta direção, esta dissertação de Mestrado traz uma inovação no sentido que está sendo proposto um sistema que permite que um robô humanoide seja inserido em salas de aulas. Trata-se de um protótipo que permite que o robô reconheça figuras geométricas planas, que pode ser estendido para outros tipos de conteúdos. O objetivo é a integração de um sistema de visão computacional em um ambiente de controle de um robô humanoide para torná-lo capaz de reconhecer figuras geométricas planas, para ser utilizada como uma ferramenta de ensino. Este sistema de visão é baseado em técnicas de Atenção Visual e utiliza uma rede neural LEGION para segmentar os objetos mais salientes da imagem e uma rede neural do tipo Multicamadas (MLP), para realizar a classificação desses objetos. Graças a este sistema de visão, o robô consegue discernir figuras sobrepostas independente do ambiente real no qual esteja inserido. Para avaliar o desempenho do sistema proposto, algumas aplicações foram desenvolvidas que envolveram a participação de crianças interagindo com o robô no reconhecimento de figuras geométricas. Embora sejam necessários uma maior número de experimentos, os resultados obtidos indicam que o sistema proposto apresenta-se como uma ferramenta alternativa, promissora e interessante, tendo sido muito bem recebida por parte dos alunos e professores.

Abstract

Increasingly there has been the introduction of new technologies in the classroom. With the help of public policies, computers connected to the internet has been used in the classes in the far reaches of our country. This has provided the knowledge reaches broader and unrestricted way to all children under development. In the last decade, besides computers, has been seen the presence in classrooms of Ipads, smart phones, owned by the students themselves and even electronic whiteboard in schools with higher purchasing power. Added to this, there is also the inclusion of robotic kits that has motivated much students in learning logical reasoning and programming, as they experience the concept: "learning through doing". The use of all these technologies aims to captivate the attention of students, encourage research and interactive learning, since the school, before exhibition, gives way to interactive teaching, that is, who has the most active student participation. In this direction, this Master's thesis brings an innovation in the sense that is being proposed a system that allows a robot humanoid is inserted into classrooms. It is a prototype that allows the robot to recognize planar geometric figures, which can be extended to other types of content. The goal is the integration of a computer vision system in a control a humanoid robot environment to make it able to recognize This has provided the knowledge gets broader and unrestricted way all children under development geometric figures, to be used as a teaching tool. This vision system is based on Visual Attention techniques and uses a neural network LEGION to target the salient objects image and a Multilayer (MLP) neural network, to perform the classification of these objects. Thanks to the vision system, the robot can distinguish independent of the actual environment in which overlapping figures is inserted. To evaluate the performance of the proposed system, some applications were developed that involved the participation of children interacting with the robot in the recognition of geometric figures. Although a larger number of experiments are needed, the results indicate that the proposed system is presented as an alternative tool, promising and interesting, and it was very well received by students and teachers.

Sumário

Resumo	v
Abstract	vii
Sumário	ix
Lista de Figuras	xi
1 Introdução	1
1.1 Objetivos	5
1.2 Motivação	5
1.3 Organização	6
2 Revisão Bibliográfica	7
2.1 Reconhecimento de padrões	7
2.2 Redes Neurais	9
2.3 Atenção Visual	13
2.3.1 Enviesamento top-down	14
2.3.2 Seleção baseada em objetos	15
2.4 Robótica Social	16
2.5 Considerações finais	18
3 Fundamentação Teórica	19
3.1 Espaço de cores	19
3.1.1 Transformações lineares	20
3.1.2 Transformações não-lineares	21
3.1.3 Espaços Híbridos	22
3.2 Segmentação de Imagens	22
3.2.1 Métodos de Segmentação	23
3.3 Atenção Visual	25
3.3.1 Mapa de Saliencia	25
3.4 Redes Neurais	28
3.4.1 <i>MultiLayer Perceptron (MLP)</i>	28
3.4.2 <i>BackPropagation</i>	32
3.4.3 Rede LEGION	34
3.4.4 Mapa SOM	37
3.5 Considerações finais	38

4 Metodologia	39
4.1 Robô NAO	39
4.1.1 Hardware	39
4.1.2 NAOqi	41
4.1.3 C++ SDK	41
4.2 Arquitetura	42
4.2.1 Descrição do Modelo	43
4.2.2 Modelo <i>bottom-up</i>	44
4.2.3 Modelo <i>top-down</i>	45
4.3 Implementação no robô NAO	46
4.3.1 Bibliotecas do NAO utilizadas	47
4.4 Considerações Finais	48
5 Experimentos	49
5.1 Cenário de testes	49
5.2 Testes com a presença do robô	50
5.3 Questionário final - Avaliação das atividades pelos alunos	55
5.4 Discussão dos resultados	59
5.4.1 Atividade 1	59
5.4.2 Atividade 2	60
5.4.3 Atividade 3	61
5.4.4 Avaliação dos Alunos	63
5.5 Considerações Finais	64
6 Conclusão e Trabalhos Futuros	65
6.1 Conclusão	65
6.2 Trabalhos Futuros	67
Referências Bibliográficas	78

Lista de Figuras

1.1	Sistema geral de reconhecimento de padrões. Fonte: (Chavez, 2002)	3
1.2	Exemplos de variações nas imagens. Fonte: (Chavez, 2002)	3
2.1	Arquitetura proposta em Wolfe (1994)	15
3.1	Cubo das cores RGB (Starna, 2009)	20
3.2	Hexágono do modelo de cores HSV. Fonte: (ENVI, 2000)	22
3.3	Duplo cone do modelo de cores HSI. Fonte: (Crabtree et al., 1984)	23
3.4	Normalização dos mapas de conspicuidades (Benicasa, 2013)	28
3.5	Morfologia do Neurônio Biológico	30
3.6	O Neurônio não-linear (Quiles, 2004)	31
3.7	Arquitetura da Rede Neural MLP (Quiles, 2004)	32
3.8	Modos do Oscilador de relaxamento: (a) Modo ativo e (b) modo excitável (Wang, 1995)	36
4.1	Robô NAO do laboratório LAR-USP	40
4.2	Como o Software do NAO funciona Robotics (2014)	42
4.3	Fluxograma do Sistema Proposto	43
4.4	Segmentação do LEGION feita pelo NAO	44
4.5	Exemplos do treinamento	45
4.6	Fluxograma de funcionamento do sistema	47
5.1	Gato feito com o Tangram	51
5.2	Aluno mostra a figura do gato para o robô NAO	52
5.3	Jogo de perguntas com as crianças	54
5.4	Com a resposta errada, o robô se agacha e fica vermelho	55
5.5	Exemplo da primeira imagem	56
5.6	Segunda imagem, um pouco mais complexa	56
5.7	Aluno responde a uma pergunta do NAO	57
5.8	O NAO levanta os braços em sinal de alegria pela boa participação do aluno	57
5.9	Alunos em grupo discutindo as falhas e acertos do NAO durante todo o experimento	58

Introdução

Recriar o comportamento humano em máquinas é uma tarefa que vem sendo estudada desde o início da terceira fase da revolução industrial, com o início da robótica e o aumento da importância e motivação para se trabalhar nessa área (Beauchamp, 1998). A Inteligência Artificial (IA) estuda e tenta reproduzir o comportamento racional, seja de animais (com organizações como os enxames) até a complexa mente humana.

A inteligência passa por fases, partindo de um conhecimento prévio até a capacidade de aprender e de tomar decisões (Luger, 2004). A IA é a arte de criar máquinas que executam comportamentos que exigem inteligência (Kurweil, 1990). E mesmo dentro da IA, temos diversos problemas para serem abordados, como aprendizado, percepção, jogos, demonstração de teoremas e diagnósticos de doenças, que envolvem diversas áreas (psicologia, matemática, computação) e algoritmos.

A partir da década de 80, foi considerado que, para evoluir, a IA necessitava de uma mudança de paradigmas (Brooks, 1985), e os conceitos da robótica saíram da pura automação dos processos industriais para o estudo e modelagem do comportamento de animais. Pouco a pouco exemplos de sistemas inteligentes foram surgindo e passaram a fazer parte do cotidiano da comunidade em geral. Jogos que aprendem e se adaptam aos movimentos do jogador, sistemas médicos que apoiam o diagnóstico de doenças a partir dos sintomas, sistemas que reconhecem rostos em fotos, são todos exemplos de avanços das pesquisas nesta área e que hoje já são muito comuns.

Dentre as aplicações que usam a IA temos o reconhecimento de padrões, como no caso de reconhecimento de pessoas a partir de extração de características da face (Gaspar, 2006). Defini-se a tarefa de reconhecimento de padrões como sendo a forma em que a máquina observa o meio, distingue padrões (entidade, objeto, processo ou evento) de interesse e toma decisões (Jain et al., 2000). O reconhecimento de padrões é uma tarefa realizada com eficiência por humanos e animais desde os primeiros dias de vida (Russel e Norvig, 2004), mas é um problema bastante complexo para um sistema computacional. Além de entretenimento, o reconhecimento pode ser usado na área médica, para identificação de anomalias fisiológicas em imagens radiográficas, na indústria para detectar padrões de falha em peças ou na defesa, como uma importante ferramenta de localização de alvos em imagens de satélite e aerofotogrametria (Haykin, 2001), por exemplo.

Um esquema geral para tratar o reconhecimento de padrões pode ser visto na Figura 1.1. A partir dos dados de entrada recebidos pelo sensor (câmera) esses dados passam pela etapa de extração de características, onde um espaço constituído de exemplos (padrões) é considerado para o mapeamento de um vetor de características. O propósito dessa extração é a redução da quantidade de dados através da observação de propriedades que distinguem os padrões de entrada. A partir desses padrões, um algoritmo de classificação pode rotular corretamente um objeto, de acordo com os dados de treinamento.

As Redes Neurais Artificiais (RNA) são utilizadas para resolver essa classificação dos objetos. As RNA se baseiam no sistema nervoso humano, imitando a arquitetura de neurônios interligados, nas quais o aprendizado ocorre por meio da adaptação dos pesos das conexões (sinapses) e do conhecimento extraído dos padrões (exemplos) apresentados a elas (Haykin, 2001).

Para que o sistema aprenda, é necessário que os padrões usados para o treinamento façam uma boa representação do problema, diminuindo a possibilidade da rede errar. Porém, os padrões na vida real apresentam algumas deformações, exemplificadas na Figura 1.2. Situações como iluminação ruim, distância do objeto para o sensor, imagem com algum tipo de ruído ou mesmo o objeto rotacionado, dificultam ainda mais a tarefa de reconhecimento automático, apesar de ser uma tarefa fácil para o cérebro humano. Como visto em (Moik, 1980; Khatchatourian e Padilha, 2008; Tejada, 2008), o uso de RNAs para se resolver o problema do reconhecimento de padrões já foi amplamente abordado.

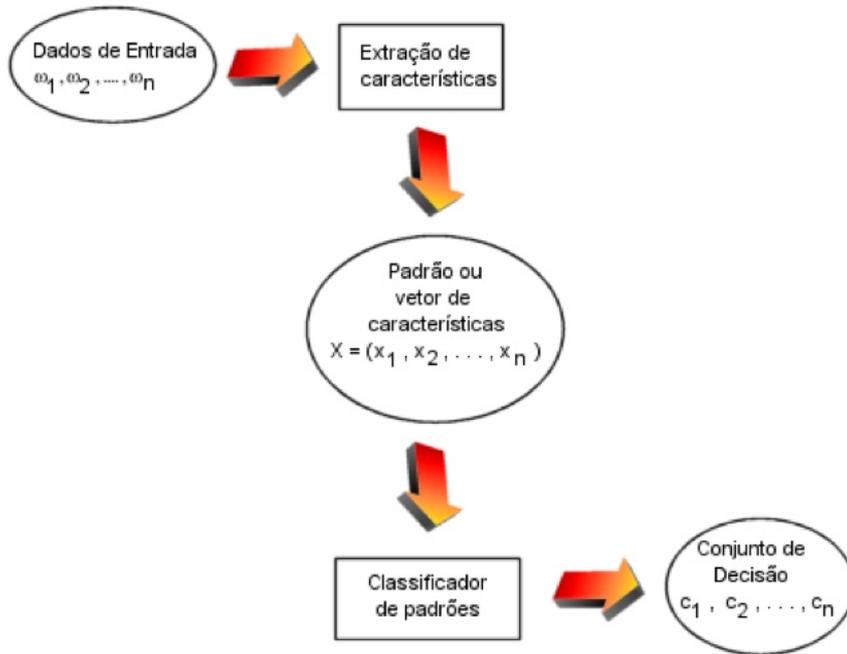


Figura 1.1: Sistema geral de reconhecimento de padrões. Fonte: (Chavez, 2002)

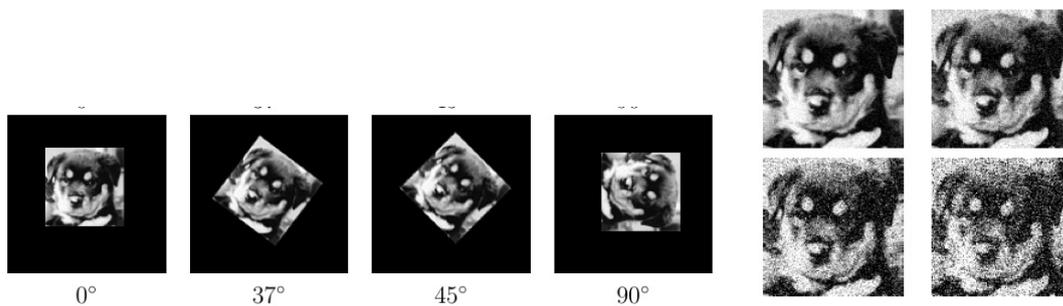


Figura 1.2: Exemplos de variações nas imagens. Fonte: (Chavez, 2002)

Se a tarefa de reconhecimento já é complexa, tratar informações em tempo real a torna um desafio ainda maior. Para diminuir a carga computacional [Shic e Scassellati \(2007\)](#), é interessante conseguir focar em um objetivo, no processo chamado de atenção visual. Esse processo é tão importante nos sistemas biológicos que chega a ser tido como determinante para preservação e perpetuação das espécies [Itti e Koch \(2001\)](#). Para tal, regiões mais relevantes ou salientes são consideradas primeiro (como um predador/presa, por exemplo), descartando-se informações menos relevantes. Um processo análogo a esse pode ser aplicado em sistemas inteligentes de visão.

De forma geral, o processo de seleção (ou atenção) visual seleciona apenas um subconjunto da informação sensorial, em uma região chamada de foco de atenção. Apenas informações que estão nessa área são consideradas, suprimindo-se o restante, sendo um eficiente mecanismo para reduzir a complexidade de tarefas como análise de uma cena, sub-dividindo em tarefas menores. Em geral os estudos apontam para dois modelos gerais:

- Um modelo baseado na localização de estímulos da cena, que guiam a atenção visual (*bottom-up*)
- Modelo baseado nas características intrínsecas do objeto, sendo dependente da tarefa (*top-down*)

mas ambos os modelos atuam na seleção de estímulos relevantes em uma cena, diminuindo a complexidade temporal para o reconhecimento de objetos.

Com esse modelo de reconhecimento, o sistema pode ser aplicado a máquinas para atuar em tempo real, como o caso de robôs. Utilizando um robô humanóide, é possível permitir que ele faça o reconhecimento de padrões e dê um *feedback* imeditado por meio de voz, luz e movimentos.

Nesta dissertação, foi utilizado um sistema para o reconhecimento de padrões, baseado em técnicas computacionais, que implementa o mecanismo da atenção visual. Este sistema foi incorporado a um robô humanóide, NAO, existente no Laboratório de Aprendizado de Robôs Móveis - LAR do ICMC, visando fornecer uma ferramenta alternativa para o ensino de figuras geométricas planas aos estudantes do Ensino Fundamental.

1.1 Objetivos

O objetivo desta dissertação de Mestrado é a integração de um sistema de visão computacional em um ambiente de controle de um robô humanoide visando torná-lo capaz de reconhecer figuras geométricas planas, para ser utilizado como uma ferramenta de ensino.

1.2 Motivação

Robôs sociais são agentes que fazem parte de um grupo de robôs ou humanos, sendo capazes de reconhecer os demais e realizar interações sociais. Para tal, esses agentes devem interpretar o ambiente e se relacionar de maneira natural, através de fala, expressões, gestos ou movimentos do corpo (Breazeal, 2002).

Existem motivações científicas e práticas para o desenvolvimento de robôs sociais. Tenta-se buscar inspiração no entendimento do comportamento social dos humanos visando aprimorar o processo de aprendizagem dos robôs, de forma que aumente a utilização destes no auxílio ao homem, em um número cada vez maior de tarefas. Neste contexto, alguns pesquisadores têm empregado robôs como ferramentas de ensino em terapias com crianças autistas (Robins et al., 2004) (Bjoorne e Balkenius, 2005).

O pesquisador Ashby (1952, 1956) já defendia que o cérebro humano trabalha por processos mecânicos e pode, pelo menos em partes, ser reproduzido em máquinas. Papert (1994) diz que a escola está no contexto de sociedade e deve viver a mesma revolução tecnológica, fazendo necessária a reflexão crítica sobre a influência da tecnologia e da interação humano-computador no desenvolvimento das pessoas, no que diz respeito a criatividade, raciocínio lógico e assimilação de disciplinas. Em (Duffy et al., 1999), o conceito de robótica social é discutido, afirmando que a sociabilidade implica em relacionamentos interativos.

Considerando estas ponderações e também o fato que os robôs humanóides atraem muita a atenção tanto de crianças como de adultos, a inserção dos mesmos em salas de aulas, auxiliando na aprendizagem, pode tornar o aprendizado mais interativo e interessante para os alunos.

As maiores dificuldades a serem tratadas consistem na qualidade da imagem obtida (por problemas de luz e sombra, por exemplo) e no espaço no qual está inserido o objeto a ser reconhecido (se existem outros objetos com a

mesma forma geométrica, cores dos objetos em volta). O uso de um sistema de atenção visual no momento da extração de características deve garantir a boa acurácia desta classificação, permitindo que o robô reconheça o objeto independentemente de sua posição na imagem. Com o robô capaz de identificar o objeto pela forma e pela cor, será implementado um módulo de reação do NAO, apontando para o objeto ou colorindo seus olhos da mesma cor reconhecida, como forma de estímulo para uma criança. Espera-se que este estudo seja o princípio de uma abordagem alternativa de técnicas de ensino na qual o uso de robôs seja considerado, principalmente em casos de crianças com algum tipo de dificuldade de aprendizagem.

1.3 Organização

Este documento está organizado da seguinte maneira. No **Capítulo 2**, é apresentada uma revisão de trabalhos sobre o tema desta dissertação, como o conceito de reconhecimento de padrões e atenção visual, bem como uma seleção de trabalhos sobre técnicas de reconhecimento e robótica educacional. No **Capítulo 3**, os principais conceitos de visão computacional são apresentados, tais como, espaço de cores, segmentação e classificação dos objetos, destacando-se os métodos utilizados nesta pesquisa, como, as redes neurais LEGION e a rede Multicamdas MLP. No **Capítulo 4**, as funcionalidades do robô humanoide, NAO, utilizado nos testes, são descritas. Todo o processo de integração entre os algoritmos do Capítulo 3 com o sistema operacional e ferramentas utilizadas do robô NAO, são também apresentados. No **Capítulo 5**, visando testar o desempenho o sistema integrado propostas, são descritas várias atividades que contam com a participação com crianças do ensino fundamental. Finalmente, no **Capítulo 6**, uma conclusão sobre os resultados obtidos é apresentada, seguida dos trabalhos futuros que poderão ser realizados dando continuidade a esta pesquisa.

Revisão Bibliográfica

A literatura mostra os avanços nas pesquisas e os problemas encontrados nas abordagens propostas para a tarefa de reconhecimento de padrões. Neste capítulo, os passos que compõem a tarefa de reconhecimento de padrões são descritos e os vários trabalhos encontrados na literatura sobre o uso de RNAs em reconhecimento de padrões, são apresentados. Além disso, conceitos e modelos de atenção visual e alguns trabalhos de robótica social são também apresentados.

2.1 Reconhecimento de padrões

Como visto no Capítulo 1, o casamento de padrões em imagens consiste na verificação de um determinado *template* (objeto padrão ou de referência) em uma imagem digital. Essa é uma das tarefas tratadas pela visão computacional, usada para fazer o registro de imagens, detecção de movimento, análise de imagens médicas e reconhecimento de objetos (Conci et al., 2008). Por padrão, entende-se que seja qualquer coisa que possa ser definida em termos quantitativos, o que inclui objetos caracterizados por sua cor, forma (foco deste Mestrado) além de outros atributos. Vale lembrar que algumas aplicações requerem que seja encontrada uma ocorrência exata do padrão, enquanto outras permitem uma ocorrência aproximada.

Segundo Gonzalez e Woods (2002), e é uma ideia compartilhada por muitos pesquisadores como (Zhang e Lu, 2004; Choksuriwong et al., 2005;

Veltkamp e Latecki, 2006), um sistema de visão computacional deve ser capaz de extrair apenas informações importantes nas imagens, fazer inferências a partir da incompletude de dados de entrada e conseguir identificar padrões com a maior independência possível em relação a mudanças de posição, orientação, tamanho e transformações geométricas.

Um sistema de reconhecimento envolve, normalmente, as etapas de aquisição de imagens, pré-processamento, segmentação, extração de características e o reconhecimento (Long et al., 2005). Assim, a maioria dos métodos propostos na literatura faz a simplificação da imagem, com a segmentação e detecção de bordas por exemplo, para em seguida, extrair as características do padrão procurado. Deve-se, porém, tomar cuidado para essa simplificação não descartar informações importantes, tornando o reconhecimento sujeito a erros.

Segundo Pedrini e Schwartz (2008); Liu e Yang (2008), uma importante propriedade na análise de imagens, mesmo quando feita por humanos, é a cor. Isso é verificável, uma vez que a cor é uma característica visualmente dominante e descritiva em uma imagem. Por isso, é o aspecto mais simples de se extrair de uma imagem (Conci et al., 2008). Assim, o reconhecimento de cores será um auxílio ao reconhecimento das formas propostas neste Mestrado.

Nos últimos anos é notório o acesso facilitado a câmeras de vídeo de alta resolução, gerando um aumento na qualidade das imagens, além do surgimento de computadores com alta capacidade de processamento, permitindo pesquisas que abordem resoluções a problemas mais complexos. Esse aumento gerou um maior interesse em pesquisas de visão computacional, gerando uma demanda por aplicações mais robustas, resistentes a variações de brilho e contraste, tolerantes às transformações afins (explicado no Capítulo 3) e insensível a ruídos.

Ressalta-se, portanto, a necessidade de métodos que possuam invariância a mudanças de orientação (rotação e translação) e escala, pois objetos semelhantes de tamanhos diferentes ou em diferentes posições devem ser considerados similares independentes dessas condições. A literatura mostra um extenso número de descritores que resolvem esse problema. Um estudo mais aprofundado sobre isso pode ser encontrado em (Sakar, 2000; Flusser e Suk, 1996; Zitova e Flusser, 2003). Algumas técnicas a serem investigadas neste trabalho são apresentadas no capítulo 3.

2.2 Redes Neurais

O uso de redes neurais artificiais na tarefa de reconhecimento de padrões têm sido amplamente cobertos pela literatura, sendo uma das técnicas mais utilizadas nesta tarefa ([Chakraborty et al., 2012](#)). Vale, portanto, destacar as vantagens do uso destas redes para entender o seu relacionamento e a sua aplicação no processo de reconhecimento de padrões. Notadamente, a tarefa de reconhecimento de padrões, dentre os domínios de aplicação, é aquela onde as RNAs possuem maior potencial ([Wang, 1995](#)).

Por sua própria natureza, o reconhecimento de padrões é uma ciência não-exata ([Bezdek et al., 1992](#)). Existem padrões bem estruturados e de fácil reconhecimento, enquanto outros padrões são não-estruturados, necessitando abordagens mais complexas para seu reconhecimento. O uso das RNAs aparece, segundo [Iyengar e Kashyap \(1991\)](#), como um bom modo de resolução pois a construção dessas redes envolve um entendimento informal do comportamento, ao invés de criar procedimentos lógicos, sendo portanto mais fácil de ser implementada

Em [Bezdek et al. \(1992\)](#) destaca-se as quatro maiores vantagens do uso de redes neurais sobre muitas outras técnicas de reconhecimento de padrões.

- adaptatividade: habilidade de se ajustar a novas informações;
- velocidade: via o paralelismo massivo;
- tolerância a falhas: capacidade de oferecer boas respostas mesmo com falta, confusão ou dados ruidosos;
- otimalidade: visto a baixa taxa de erros em sistemas de classificação.

De modo geral, as redes neurais são um método de modelamento altamente recomendável para se lidar com sistemas abertos ou mais complexos, pouco entendidos e que não podem ser adequadamente descritos por um conjunto de regras ou equações. Como cores e formas geométricas são padrões bem conhecidos para se treinar a rede, sua escolha é justificável pela facilidade de implementação e rapidez de resposta, uma vez que a rede esteja treinada.

Conforme [Egmont-Petersen et al. \(2002\)](#), diversos trabalhos têm utilizado RNAs no processamento de imagens, diretamente no processo de segmentação ou como ferramenta auxiliar. Alguns destes são apresentados a seguir:

Em [Ong et al. \(2002\)](#) é proposto um método de segmentação de imagens coloridas baseado em classificação de cores usando mapas auto-organizáveis

(SOM - *Self-Organizing Maps*). De forma não supervisionada, uma rede neural SOM 2D era utilizada para fazer um mapeamento e clusterização das cores nas imagens de treinamento e uma outra rede 1D extraía elementos representativos destes *clusters*. Essa segunda rede podia sofrer agrupamentos e divisões, para aglomerar cores muito próximas em um *cluster* ou dividir um outro em mais cores. O espaço de cores (tema discutido no Capítulo 3) utilizado no processo de segmentação foi o $L^*u^*v^*$. Apesar de apresentar bons resultados, este trabalho não foi desenvolvido para aplicações de tempo real, sendo necessários ajustes para esta finalidade.

Um método de pesquisa similar foi utilizada por Papamarkos (1999), porém utilizando apenas uma rede SOM e mais algumas características, como média da vizinhança e entropia no espaço RGB. Em Papamarkos et al. (2002) o sistema foi melhorado com a inclusão de uma rede PCA (*Principal Component Analysis* - Análise de Componentes Principais) para diminuir a correlação das características apresentadas à rede SOM.

Os processo proposto por Papamarkos et al. (2002) apresentam bons resultados no processo de redução de cores com o espaço RGB, mas não são métodos aplicáveis para atuar em tempo real, pois a rede precisa ser treinada para cada imagem que sofrerá redução de cores, limitando a utilização destas propostas.

Em Simoes e Costa (2000a) e Simoes e Costa (2000b) é usada uma rede neural MLP treinada via o algoritmo *backpropagation* (BK) no processo de segmentação para o domínio de futebol de robôs. O espaço de cores utilizado foi o RGB, onde os três componentes de cada ponto da imagem eram as entradas da rede. 7 classes de cores foram estabelecidas: laranja, verde, rosa, branco, amarelo, cinza e preto.

Os resultados mostraram que a classificação de cores usando MLP é uma boa alternativa para o domínio aplicado, pois separa o espaço de cores por hiperplanos, diferente da limiarização tradicional dos métodos de segmentação clássicos. Além disso, testes comparativos em Simoes e Costa (2001) mostraram o espaço RGB com melhores resultados do que o espaço HSV e YUV. Porém, as imagens obtidas eram através de uma câmera no teto, o que diminuía as variações de luz e sombra presentes no ambiente, não sendo a técnica robusta em outras condições.

Já em Tsai e Tseng (2012), foi preferido o uso de um espaço HSL para fazer o reconhecimento de cores no ambiente de futebol de robôs, que obteve uma acurácia melhor para materiais brilhantes, mesmo em diferentes condições

de iluminação. Além disso, foi proposto um algoritmo que diminui as falhas apresentadas no espaço de cores RGB.

No caso de [Pomerleau \(1995\)](#) foi utilizada uma Rede Neural para controlar um veículo de forma autônoma (ALVINN). A imagem foi pré-processada para diminuir a quantidade de dados a serem tratados pela rede, e sua saída representava a direção que o carro deveria tomar. Esse trabalho possibilitou que o veículo navegasse por rodovias a mais de 88 km/h, mostrando que uma rede bem treinada era eficiente para o controle de um robô, mesmo em ambientes desconhecidos.

Em [Blank e Ross \(1997\)](#) são usadas as RNAs como entrada de um sistema fuzzy que controlava um robô na busca por uma bola, mapeando as imagens de baixo nível em entradas de auto nível para o controlador. Este trabalho já mostrou um certo tratamento de cores, já que a bola tinha uma coloração bem forte, mas falhou em alguns aspectos importantes, como o não tratamento do caso em que a bola se encontrava fora dos quadrantes de visão inicial do robô.

[Waldherr et al. \(2000\)](#) também é trabalhado o reconhecimento de cores para criar uma interface de reconhecimento de gestos para o controle de um robô. A partir da cor da camisa e do rosto, foi realizado o *tracking* do humano, reconhecendo-o para interpretar os gestos. Assim, a movimentação dos braços proporcionou o envio de comandos do tipo siga, pare, etc. Vale ressaltar que, no início do processo, a pessoa deveria ficar no centro de visão do robô. De modo parecido, em [Tzafestas et al. \(2009\)](#) foi classificado, também por meio de uma MLP, diversos gestos da mão para o controle de um robô móvel, obtendo uma acurácia de 98.5%.

Em [Quiles \(2004\)](#), um sistema de visão foi criado para reconhecer objetos de cor e forma determinada, para o controle de robôs móveis. Este trabalho mostrou um ótimo resultado para o reconhecimento de cores, permitindo que o robô seguisse muito bem as cores determinadas (azul, vermelha e amarela) com o uso de Redes Neurais e aplicações de campos potenciais. Entretanto, o uso de apenas uma rede se mostrou ineficiente quando a tarefa era reconhecer cores e formas. Isso acontecia porque na etapa de pré-processamento, o reconhecimento de cores acabava por gerar uma forma deformada do objeto, dependendo das condições de iluminação, escala, etc. Este trabalho mostrou a importância de se utilizar as técnicas invariantes no tratamento de formas, para evitar esse tipo de erro.

[Quiles et al. \(2006\)](#) utilizaram uma PCNN (Pulse Coupled Neural Network) ([Lindblad, 2005](#)) para fazer um modelo aplicado diretamente a segmentação.

Pulsos sincronizados representavam os objetos da imagem, e grupos de neurônios fora de fase representavam os objetos distintos. O modelo proposto conseguiu separar objetos não-linearmente separáveis, tarefa bastante complicada para os demais algoritmos de segmentação.

Em [Fiaz e Ijaz \(2010\)](#), uma MLP de três camadas foi utilizada para o rastreamento de atividades humanas em sistemas de segurança. Na etapa de segmentação, características de movimento eram extraídas, de forma a criar os padrões de comportamento a serem reconhecidos. Os comportamentos eram: nenhuma atividade, levantar-se, sentar-se, curvar-se, andar e correr. A MLP teve uma alta acurácia no reconhecimento destes comportamentos, atingindo 98%, por exemplo, nos momentos do vídeo em que a pessoa estava sentando ou levantando. O comportamento de corrida foi onde a rede mais errou, tendo uma acurácia de 81% em 1010 casos apresentados durante a avaliação.

Zaman ([Zaman et al., 2011](#)) comparou diferentes algoritmos de redes neurais para um sistema de controle de tráfego distribuído (*Distributed Traffic Light Control System*), que tem como objetivo diminuir o risco de congestionamentos. Escolhidos para otimizar o reconhecimento dos diferentes veículos, foram testados o *backpropagation*, o *Fuzzy Learning Vector Quantization* (FLVQ) e o *Fuzzy Learning Vector Quantization Particle Swarm Optimization* (FLVQ-PSO). A média da acurácia obtida pelo algoritmo *backpropagation* foi de 99.51%, contra 90.71% da FLVQ e 74.76% da FLVQ-PSO. A superioridade também foi demonstrada no tempo de convergência, pois o *backpropagation* precisou de 45 épocas para convergir, enquanto os demais convergiram após 100 épocas.

Ainda sobre comparação de métodos, em [Chakraborty et al. \(2012\)](#) é implementada uma rede MLP com o algoritmo Levenberg-Marquardt ([Chatterjee et al., 2002](#)) de otimização, melhorando o *backpropagation*, durante o treinamento da rede. Essa abordagem mostrou uma superioridade quando comparada à tradicional no tempo de busca da rede. Foi demonstrado também uma invariância a escala e rotação. O pesquisador conclui sua pesquisa afirmando que essa redução pode ajudar em aplicações de robótica que trabalham em aplicações em tempo real.

A boa classificação da MLP também foi obtida em ([Spasojevic et al., 2012](#)), onde a rede foi usada para classificar cubos, pirâmides e cilindros. As imagens eram segmentadas para limpar o fundo e, depois de feita a extração das bordas, eram classificadas pela rede neural. Para imagens não utilizadas

no treinamento, a rede classificou de forma errônea apenas uma vez com os cubos e piramedes e teve 100% de acerto para os cilindros.

2.3 Atenção Visual

A atenção visual tem sido muito pesquisada nos últimos anos. Os vários modelos propostos variam de acordo com a modelagem e a realização e comparação dos testes. Em [Tsotos \(2011\)](#) e [Borji e Itti \(2013\)](#) é apresentada uma revisão detalhada desses trabalhos relacionados à essa abordagem. No presente trabalho, vamos considerar os trabalhos baseados nas seguintes hipóteses:

- **Mapa de Saliência:** A partir da extração de características específicas são calculados mapas da integração dos estímulos locais. A partir deste mapa é possível descobrir qual característica é mais saliente.
- **Atenção Emergente:** Interações competitivas entre uma grande quantidade de neurônios geram a definição da região de maior atenção, enviesado por um mecanismo *top-down*.
- **Correlação Temporal:** Sistemas compostos por grupos de neurônios com conexões excitatórias e inibitórias, cujo dinamismo é regido por equações diferenciais

Os mapas de saliência, segundo [Itti e Koch \(2001\)](#), são um processo rápido de busca pelo local de maior atenção e são dirigidos principalmente pela forma *bottom-up*. Estímulos salientes, acima de um *threshold*, serão destacados automaticamente na cena, o que faz essa abordagem sensível ao contexto ([Ogawa e Komatsu, 2004](#)).

Através do mapa de saliência proposto em [Koch e Ullman \(1985\)](#), é possível direcionar a atenção para pontos de maior interesse na imagem. É uma proposta de atenção seletiva: cada característica primitiva (orientação, cor ou intensidade) é calculada e representada por um mapa representativo dessa característica. Esses mapas são criados, paralelamente, e representam as regiões de maior contraste de cada um dos atributos. Esses mapas são combinados em um único outro, que representa a conspicuidade de toda a cena visual.

Uma rede *Winner-Take-All* é utilizada para encontrar a região do ponto mais saliente. Essa seleção é encaminhada para a representação central. De forma

sequencial, a rede WTA indicará o segundo ponto de saliência, repetindo o processo até se esgotar as regiões salientes da cena.

O modelo proposto por [Itti et al. \(1998\)](#) baseia-se diretamente desta categorização apresentada. As características visuais passam primeiro por uma filtragem composta por 8 escalas espaciais, enfatizando o contraste local de cada atributo, usando o cálculo da diferença de centro-vizinhança. No momento da detecção pela WTA existe o mecanismo de inibição de retorno, que impede que uma região já determinada como saliente seja novamente visitada. Vários outros trabalhos estenderam esse modelo, é o caso dos trabalhos de [Walther et al. \(2002\)](#), [Rutishauser et al. \(2004\)](#), [Walther et al. \(2005\)](#) e [Harel et al. \(2006\)](#).

2.3.1 Enviesamento top-down

Nos modelos apresentados até aqui nenhuma informação sobre o que se quer ser reconhecido é considerada. Porém, essas informações *top-down*, de acordo com [Desimone e Duncan \(1995\)](#), influenciam a atenção, conforme conceitos neurobiológicos. Em [Borji e Itti \(2013\)](#), os autores defenderam que ambas os modelos (*top-down* e *bottom-up* não são mutuamente exclusivas, podendo-se direcionar a atenção voluntariamente para os objetos previamente conhecidos.

Em [Wolfe \(1994\)](#), o enviesamento *top-down* é realizado através de atribuição de pesos associados ao mapa de saliência, partindo-se da premissa de que alguma característica do que se quer ser encontrado é conhecida (por exemplo uma bola vermelha em meio a bolas azuis). A arquitetura proposta pode ser vista na [2.1](#). Essa associação de pesos, entretanto, nem sempre pode acelerar a busca por um alvo ([Elazary e Itti, 2010](#)), podendo ou não afetar o desempenho do modelo.

Essa abordagem, também utilizada em outros trabalhos na literatura, não é eficiente quando as características procuradas são compartilhadas tanto por alvos como pelos distratores. Na tentativa de resolver essa questão, em [Elazary e Itti \(2010\)](#), foi proposta a utilização de um conjunto de características na geração do mapa e no treinamento do classificador, utilizando funções de probabilidades. As informações aprendidas durante o treinamento são utilizadas para guiar a atenção visual, devolvendo uma lista de possíveis objetos procurados.

[Borji et al. \(2011\)](#) tentaram aumentar a taxa de acerto de alvos incorporando custos de processamentos para cada mapa de saliência,

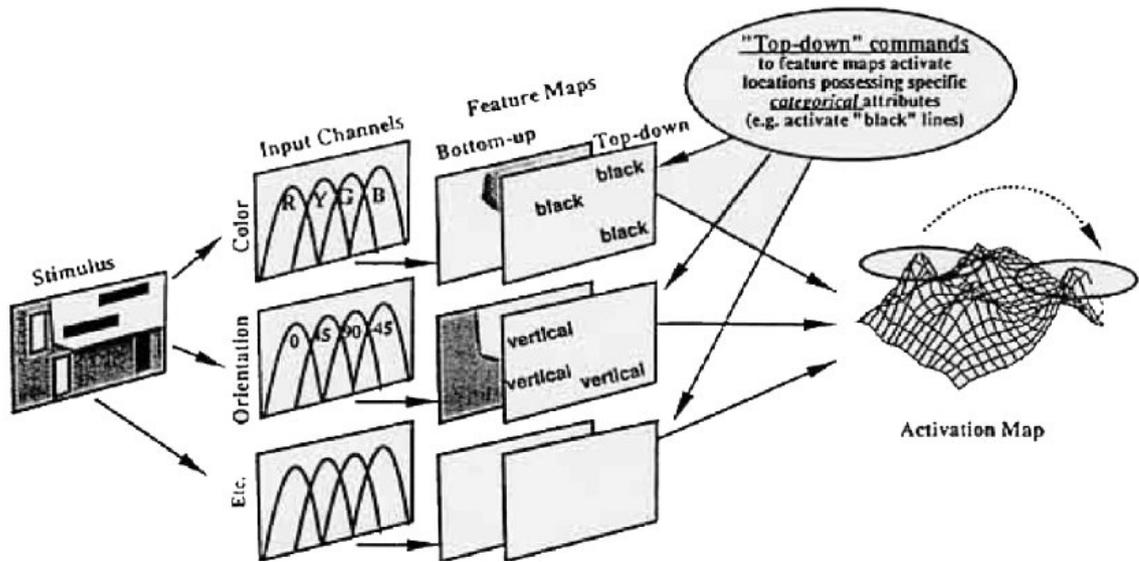


Figura 2.1: Arquitetura proposta em Wolfe (1994)

buscando uma maior taxa de detecção por um menor custo. Para tal, eles revisaram o modelo de Itti et al. (1998) para selecionar um determinado nível da pirâmide gaussiana, conseguindo inibir escalas menos importantes, alcançando o objetivo inicial da pesquisa.

2.3.2 Seleção baseada em objetos

Os trabalhos apresentados até agora se baseiam em uma busca a partir da cena (*space-based*). Porém, no sistema visual de primatas, é empregado um mecanismo de seleção baseado em objetos (*object-based*) (Benicasa et al., 2013). Uma característica fundamental da atenção é integrar características correlacionadas aos objetos (Wang, 2005). Essa idéia está diretamente ligada à teoria da correlação temporal von der Malsburg (1981), e é obtida através da sincronização e dessincronização das atividades de neurônios espacialmente distribuídos. Ou seja, osciladores representam um conjunto de características, fazendo com que cada objeto seja representado por um conjunto de osciladores síncronos. Assim, os neurônios que representam um objeto estão sincronizados, enquanto os neurônios de outros objetos não estão.

A partir dessa correlação oscilatória foi proposto a rede neural LEGION (*Locally Excitatory Globally Inhibitory Oscillator Network*) Wang (1995), rede de osciladores localmente excitatórios e globalmente inibitórios possibilitando, entre outras áreas, a segmentação de imagens e seleção de objetos seguindo a

metodologia descrita. Já em [Campbell et al. \(1999\)](#) esse modelo foi modificado, substituindo-se os osciladores por neurônios do tipo I& F (Integra e Dispara) melhorando a velocidade de sincronização e dessincronização dos grupos de objetos.

[Kazanovich e Borisyuk \(2002\)](#) mostraram a detecção de objetos simples de forma consecutiva, utilizando a rede LEGION, utilizando um oscilador central junto com os conjuntos de osciladores que representam os objetos da imagem. Isso aumenta a amplitude dos objetos sincronizados, criando uma maior diferença e possibilitando uma melhor definição do objeto a ser detectado. [Quiles et al. \(2011\)](#), propuseram um modelo de atenção visual baseado tanto no mapa de saliência quanto na correlação oscilatória, utilizando as redes LEGION e um mecanismo de inibição de retorno, sendo o primeiro modelo capaz de selecionar objetos em cenas reais. [Benicasa et al. \(2013\)](#) incrementaram o modelo anteriormente citado, incorporando o processo de enviesamento *top-down*. Estes dois últimos trabalhos mencionados, juntamente com aquele encontrado em [Benicasa \(2013\)](#), foram utilizados como base para esta Dissertação de Mestrado. Toda a fundamentação teórica utilizada será demonstrada no capítulo 3.

2.4 Robótica Social

Pelo fato do sistema a ser desenvolvido utilizar um robô humanoide, nesta seção alguns trabalhos, que utilizaram robôs visando o aprendizado de crianças e a repetição de movimentos, são apresentados.

Em um estudo de interação social feito por [Tapus et al. \(2012\)](#), um experimento com quatro crianças foi realizado, no qual a interação com o robô humanoide, NAO, e com um humano foi comparada. Duas dessas crianças não mostraram qualquer efeito com a presença do robô, porém, as outras duas prestaram mais atenção ao robô do que ao humano, sendo que uma delas mostrou uma grande interação com o robô. Já em outro teste, apresentado em [Csala et al. \(2012\)](#), crianças que são forçadas a ficar em pequenas caixas estéreis, de 2x3 metros, devido a uma cirurgia, tiveram respostas bem positivas, no relacionamento com o robô NAO, que as animava e convidava para fazer alguns exercícios.

[Kimberlee et al. \(2013\)](#) investigaram o uso de robôs para aumentar a comunicação e a atenção de adolescentes autistas. Para tal, adolescentes com autismo eram colocados para jogar, juntamente com três outros adolescentes

com outros distúrbios, recrutados de escolas de pessoas com necessidades especiais. Os testes foram feitos em três dias, não consecutivos, os adolescentes autistas jogavam de três diferentes formas: a) com um robô humanóide, b) com uma Smart Board e c) com cartas. Apesar de apresentarem comportamentos individualistas nos três modos, o comportamento repetitivo foi diminuído quando foi utilizado tanto o robô como a Smart Board, mostrando que é possível o uso de robôs para ajudar no aprendizado. Além disso, a pesquisa do estado-da-arte apresentada em [Shamsuddin et al. \(2012\)](#) mostra que o uso dos robôs vêm sendo testado por muitos pesquisadores, com bons resultados.

[Suay e Chernova \(2011\)](#), utilizaram o sensor Kinect junto com um humanoide, NAO, para permitir que o robô imitasse o movimento dos braços do usuário, tivesse seu caminhar ordenado e o olhar do robô fosse direcionado pela posição da mão direita para ordem de deslocamento. Estas funcionalidades foram obtidas pelo estabelecimento de alguns modos de controle com relação à posição de alguns membros do corpo do usuário.

Já em [Veltrop \(2012\)](#), o uso da fusão de vários sensores para teleoperar um humanoide NAO foi apresentado. Para tal, foram utilizados o sensor Kinect para estimação das configurações espaciais do corpo do usuário; um controle Wiimote para melhor estimação das configurações da mão/punho e ativação de determinados comportamentos no robô; uma esteira para permitir o robô caminhar conforme o caminhar do usuário na esteira evitando, assim, que o mesmo saísse do campo de visão do Kinect; e um head-mounted display servindo como a visão do robô no usuário.

[Koenemann e Bennewitz \(2012\)](#) utilizaram uma vestimenta de sensores inerciais no corpo de um usuário a fim de capturar as configurações das junções de seu corpo e mapeá-las para as junções do robô em tempo-real. Também, foi desenvolvido um módulo de balanço que permitisse o robô se comportar de forma estável durante a imitação. O robô humanoide desempenhou a imitação adequadamente durante a interação, permitindo também a estável imitação do movimento de ambas as pernas sem cair no chão.

Em Israel, [Verner et al. \(2012\)](#) utilizaram-se de diversos robôs para aprimorar a experiência de visitantes a um museu de tecnologia. Dentre eles, o robô NAO para interagir com o público explicando algumas seções do museu e respondendo algumas perguntas pré-programadas dos visitantes. Algumas das conclusões que podemos destacar foi o aumento do número de

visitantes após a implementação dos robôs; o aumento do interesse das pessoas - medido pela média de perguntas por minuto; e o quanto as pessoas memorizavam em relação ao que foi ensinado, medido por um questionário ao final da visita. Todas essas características foram significativamente melhoradas com o apoio dos robôs.

Smolar et al. (2011) estudaram o processo cognitivo do próprio robô, ou seja, a forma de aprendizado por meio de sentidos como visão, audição, etc. Já Tanaka e Ghosh (2011) inverteram o papel do NAO no ensino. Nessa pesquisa, eles programaram o robô para participar das aulas de inglês para crianças japonesas, na universidade de Tsukuba, durante 3 dias e monitoraram as interações. O robô participava de todas as atividades junto com as crianças, como cantar cantigas em uma roda e responder perguntas de cores e animais feitas pela professora. No primeiro e segundo dia, ele foi programado para responder todas as respostas corretas e no terceiro dia para errar todas as respostas. Dessa forma, no último dia as crianças advertiam o robô e explicavam o porquê de suas respostas estarem equivocadas. Os autores concluíram que as crianças criavam maneiras particulares de explicar as coisas ao robô e, com isso, potencializavam seus próprios processos cognitivos. Mediram, também, o número de interação por minuto e descobriram que o número de respostas corretas dadas pelas crianças, tanto individual quanto coletivamente, aumentavam proporcionalmente com a interação.

2.5 Considerações finais

Neste Capítulo, foi apresentada uma revisão dos trabalhos sobre os métodos a serem utilizados nesta pesquisa. A tarefa de reconhecimento de padrões é extensamente pesquisada, ainda mais com o acesso facilitado à boas câmeras, o que gera maior qualidade de fotos e maior percepção de detalhes, aumentando a necessidade de técnicas eficientes. Além disso, aplicações em tempo-real de visão computacional também são cada vez mais comuns, e exigem precisão e alta tolerância a ruídos e as transformações geométricas. A escolha da rede MLP e LEGION para a abordagem proposta é justificada pelos bons resultados que estes métodos demonstraram nos diversos trabalhos aqui descritos. No próximo Capítulo, detalhes sobre estas redes serão apresentados, bem como um estudo do espaço de cores a ser utilizado e as formas de segmentação possíveis a serem abordadas.

Fundamentação Teórica

No capítulo anterior, foram apresentados trabalhos relacionados a esta pesquisa, discutido suas contribuições e lacunas. Neste capítulo são apresentados os métodos de visão computacional e o modelo de rede neural artificial escolhidos para serem implementados neste trabalho. Na Seção 3.1, é feito um estudo sobre os diferentes espaços de cores e exemplos das transformações que são necessárias para se mudar de um espaço de cores para o outro. Na Seção 3.2, é descrita sobre uma das tarefas mais importantes do reconhecimento de padrões, que é a segmentação de cores. Os métodos de Atenção Visual, como o Mapa de Saliências e inibição de retorno são discutidos na Seção 3.3, os classificadores, as RNAs MLP e RNP, são apresentados na Seção 3.4. A Seção 3.5 apresenta as conclusões finais deste capítulo.

3.1 Espaço de cores

A percepção de cores pelos humanos é uma combinação de três estímulos de cores primárias: vermelho, verde e azul. Esse espaço de cores é o chamado espaço RGB (*Red Green Blue*), porém ele não é considerado um bom espaço para a realização do processo de segmentação da imagem, pois existe uma alta correlação entre cada um dos componentes, dificultando a representação de cores nessa forma de representação (Quiles, 2004). Isso é facilmente observado quando ocorre alguma variação de intensidade da luz,

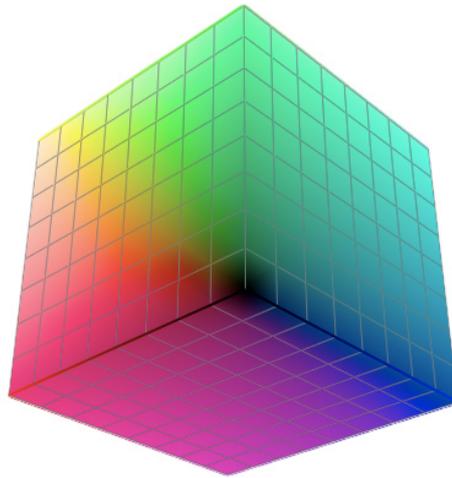


Figura 3.1: Cubo das cores RGB (Starna, 2009)

pois os três componentes têm seus valores alterados. Porém, os elementos RGB permitem a obtenção dos atributos de cromaticidade (*hue*), brilho e saturação, que são geralmente utilizados para realizar a distinção de cores.

Porém, através do espaço RGB, através do uso de transformações lineares e não-lineares, podemos gerar diversas outras representações de cores, como o HSI (*Hue, Saturation and Intensity* ou Matiz, saturação e intensidade), YUV (Onde Y é a componente de luminância e e UV são componentes de cromaticidade) e o L*a*b* (onde L indica brilho e ab são coordenadas cromáticas). Na revisão bibliográfica apresentada por Cheng et al. (2001), nenhuma delas é consagradamente melhor, sendo sua escolha dependente da aplicação.

3.1.1 Transformações lineares

Usar transformações lineares para obtenção de espaços de cores geralmente é uma tarefa de custo computacional reduzido, podendo ser utilizado em tempo real. É o caso, por exemplo, da YUV, utilizada na transmissão de sinal de televisão colorida (padrão europeu). A correlação dos componentes, entretanto, continua muito alta, mas é melhor se comparado à RGB. A transformação da YUV é dada por:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.437 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3.1)$$

3.1.2 Transformações não-lineares

Quando há necessidade de decompor cada ponto da imagem nos atributos com brilho, saturação e cor, existe a necessidade de aplicar as transformações não-lineares no espaço RGB. Esses espaços gerados (como o HSI e o L*a*b) são geralmente mais indicados para aplicar a separação de cores. Sistemas que utilizam esses espaços são mais robustos a variação de luminosidade e sombras, mas apresentam instabilidade quando aplicados a pontos com baixa saturação (Quiles, 2004). Além disso, essas transformações acarretam um maior tempo computacional, inviabilizando seu uso em aplicações de tempo real. A transformação do sistema RGB para o sistema HSI é dada por:

$$I = \frac{1}{3}(R + G + B) \quad (3.2)$$

$$S = 1 - \frac{3}{(R + G + B)}[\min(R, G, B)] \quad (3.3)$$

$$H = \cos^{-1} \left(\frac{\frac{1}{2}(R - G) + (R - B)}{[(R - G)^2 + (R - B)(G - B)]^{\frac{1}{2}}} \right) \quad (3.4)$$

Sendo MAX e MIN os valores máximo e mínimo, respectivamente dos valores (R, G, B), a transformação do sistema RGB para o HSV é dada por:

$$H = \begin{cases} 60 * \frac{G - B}{MAX - MIN} + 0 \rightarrow \text{if}(MAX = R) \text{e}(G \geq B) \\ 60 * \frac{G - B}{MAX - MIN} + 360 \rightarrow \text{if}(MAX = R) \text{e}(G < B) \\ 60 * \frac{B - R}{MAX - MIN} + 120 \rightarrow \text{if}(MAX = G) \\ 60 * \frac{R - G}{MAX - MIN} + 240 \rightarrow \text{if}(MAX = B) // \end{cases} \quad (3.5)$$

$$S = \frac{MAX - MIN}{MAX} \quad (3.6)$$

$$V = MAX \quad (3.7)$$

O modelo de cores HSV é representado por um hexágono, mostrado na

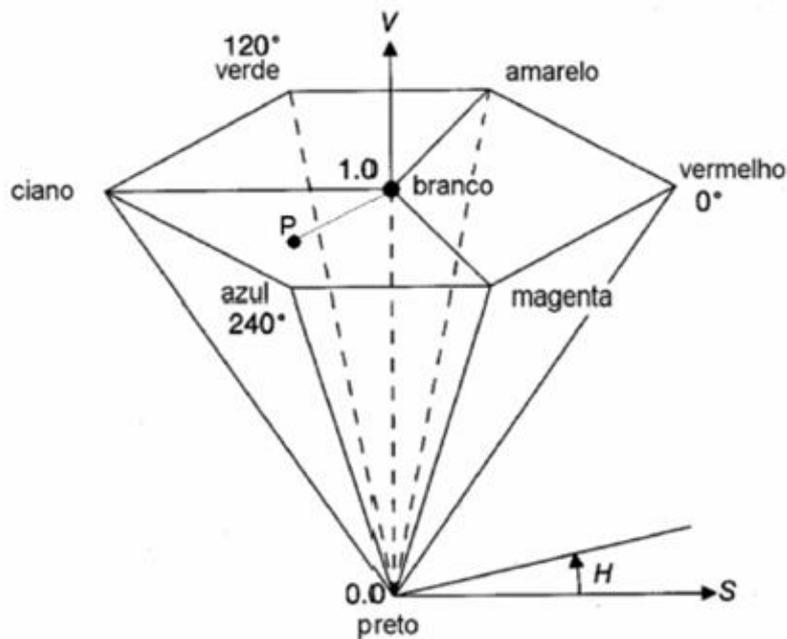


Figura 3.2: Hexágono do modelo de cores HSV. Fonte: (ENVI, 2000)

Figura 3.2 e o HSI é representado por um duplo cone, como na Figura 3.3.

3.1.3 Espaços Híbridos

Os espaços híbridos são compostos por um conjunto de elementos (como o R, G, B) obtidos através de vários espaços de cores. Com isso busca-se obter a menor correlação com o menor tempo computacional possível necessário para obtenção desses elementos, buscando obter o melhor custo/benefício entre os espaços.

3.2 Segmentação de Imagens

Um dos primeiros passos a serem realizados no reconhecimento de padrões é a segmentação de imagens, componente crucial e essencial considerado uma tarefa de difícil realização e responsável pela determinação do resultado final do trabalho (Cheng et al., 2001). Cada ponto da imagem é classificado como pertencente ou não a um determinado objeto da cena analisada (Jahne, 1997), onde cada segmento compõe uma região que é homogeneizada a fim de se destacar (separar) os objetos da imagem (Nalwa, 1993).

É também na fase de segmentação que removemos as características invariantes dos objetos, para que estes possam ser determinados

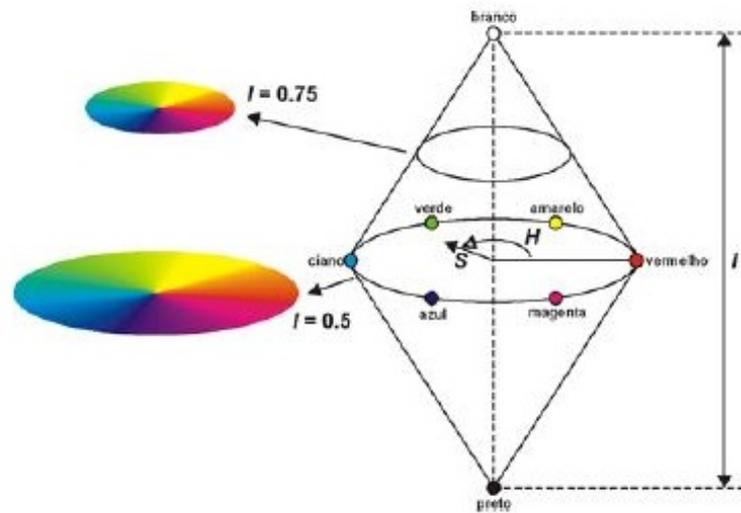


Figura 3.3: Duplo cone do modelo de cores HSI. Fonte: (Crabtree et al., 1984)

corretamente na etapa de classificação.

3.2.1 Métodos de Segmentação

A segmentação de imagens monocromáticas é geralmente baseada na homogeneidade dos níveis de cinza presentes nas imagens (Gonzalez e Woods, 2002). Nesse tipo de segmentação, a separação é realizada quando ocorrem grandes variações nos tons de cinza. Destacam-se, portanto, a detecção de bordas e linhas dos objetos presentes nas imagens.

A segmentação baseada em similaridades visa aglomerar as regiões que possuem características semelhantes, destacando-se a limiaridade e o crescimento e fusão de regiões.

A segmentação de imagens coloridas é caracterizada como expansão dos métodos acima, acrescidos de representações do espaço de cores (Cheng et al., 2001). Pode ser dividida em quatro classes distintas (Skarbek e Koschan, 1994):

- Segmentação baseada em ponto [*Pixel based*];
- Segmentação baseada em região [*Area based*];
- Segmentação baseada em bordas [*Edge based*];
- Segmentação baseada em representação física [*Physics based*];

Na detecção de bordas apenas a diferença da intensidade de luz é verificada. Isso funciona bem para imagens monocromáticas, mas para imagens coloridas, onde temos mais informações nas bordas, a tarefa se torna impossível (Cheng et al., 2001). Por exemplo uma borda com mesma intensidade luminosa mas com cores diferentes pode não ser detectada. A técnica baseada em modelos físicos descrevem propriedades da interação da luz com o material do objeto na imagem. Isso solucionaria problemas causados pelos reflexos e sombras (Lucchese e Mitra, 2001). Contudo, segundo Cheng et al. (2001), esses algoritmos só são eficientes quando conhecidas e as propriedades de reflexão dos objetos presentes na imagem, limitando a utilização dessa abordagem. Assim, serão considerados para este trabalho a segmentação baseada em pontos e em regiões.

Na segmentação baseada em ponto destaca-se a limiarização (*Histogram Thresholding*), onde é realizado um estudo do histograma da imagem para se estabelecer limiares de separação a cada um dos eixos que representam o espaço de cores, determinando-se a região de separação.

Também destaca-se nessa técnica a clusterização, que objetiva a criação de classes de cores distintas de forma que os pontos sejam classificados de acordo com estes clusters.

Nesta forma de segmentação, nenhuma informação adicional, como nível de entropia, textura ou informações espaciais, ficando restrito ao um sistema de classificação de cores.

Os métodos de segmentação baseados em regiões incluem técnicas de crescimento, divisão e agrupamento de regiões com objetivo de agrupar os pontos em regiões homogêneas. Esta técnica é largamente utilizada em segmentação de imagens coloridas, pois considera tanto informações provenientes de cor quanto as informações de textura, entropia, etc. (Cheng et al., 2001).

A técnica de crescimento deste método caracteriza-se pela expansão em determinadas regiões da imagem a partir de um ponto inicial, chamado semente. Os pontos vizinhos são agrupados a esta região, considerando-se propriedades como textura, cor, entre outros.

No processo de divisão, a segmentação parte da imagem inteira, separando as regiões consideradas heterogêneas. O agrupamento combina as técnicas de crescimento e divisão, unindo regiões vizinhas similares a fim de criar regiões homogêneas tão grandes quanto possível.

3.3 Atenção Visual

Uma vez definido o espaço de cores e o método de segmentação, nos direcionamos para as definições do processo de atenção visual. [Benicasa \(2013\)](#) propôs um modelo de atenção visual que cria os Mapas de Saliência a partir das informações primitivas da cena (enviesamento *bottom-up*) e da reposta de uma rede MLP (enviesamento *top-down*), segmentando a imagem por meio de uma rede LEGION. Esse conceito foi reproduzido nesta Dissertação, pois trata-se de uma busca no ambiente por objetos conhecidos (figuras geométricas).

3.3.1 Mapa de Saliencia

Esta pesquisa baseia-se no fato das células do córtex cerebral dos mamíferos organizam-se de forma estruturada, resultado em macro regiões do cérebro capacitadas em processamentos específicos de tarefas como linguagem, controle e visão ([Kohonen, 2001](#)). Computacionalmente, características primitivas da cena formam mapas diferentes de saliência, por isso o mapa precisa ser composto em um único, onde as regiões disputam pela atenção deixando uma única vencedora, gerando um mapa independente das características. Nesta dissertação, as características de cores e orientação são utilizadas na criação dos mapas, seguindo o modelo de [Benicasa \(2013\)](#).

De modo geral, o modelo é alimentado por um imagem e são extraídas as suas características de cor e orientação. As informações do canal I , demonstradas na Equação (3.2), são usadas para a normalização dos canais r , g e b , inibindo regiões cuja luminosidade (saliência) seja muito pequena, normalizando a região a qual I for maior do que $1/10$ do seu valor máximo sobre toda a imagem. Em seguida, quatro canais de cores são criados:

$$R = \frac{r - (g + b)}{2} \quad (3.8)$$

$$G = \frac{g - (r + b)}{2} \quad (3.9)$$

$$B = \frac{b - (r + g)}{2} \quad (3.10)$$

$$Y = (r + g)/2 - |r - g|/2 - b \quad (3.11)$$

Valores negativos de R , G , B e Y são atribuídos zero. Porém, a extração apenas desses canais de cores não retira todos os ruídos e informações indesejáveis de uma imagem. Para resolver esse problema foi utilizada a técnica da Pirâmide Gaussiana de [Burt e Adelson \(1983\)](#).

Nove níveis são considerados na criação dessa pirâmide, cujos os níveis mais altos representam convoluções das níveis anteriores, além de um processamento de filtro passa-baixo e de sub-amostragem. No trabalho de [Benicasa \(2013\)](#), foi considerado um filtro Gaussiano com dimensões de 5×5 pixels, e essa abordagem destaca as saliências de uma imagem, ao mesmo tempo que inibe as demais regiões, retirando do mapa ruídos. Esse tamanho do filtro foi mantido nesta pesquisa, pois também obteve bons resultados.

A imagem de entrada é inicialmente representada pela matriz G_0 , representando o nível zero da pirâmide, composta pelas linhas e colunas (x, y) , que representam uma coordenada da imagem. A pirâmide pode ser recursivamente definida como:

$$G_\sigma(x, y) = \sum_{m=-2}^2 \sum_{n=-2}^2 w(m+2, n+2) G(x, y) \quad (3.12)$$

$$G_\sigma(x, y) = \sum_{m=-2}^2 \sum_{n=-2}^2 w(m+2, n+2) G_{\sigma-1}(2x+m, 2y+n) \quad (3.13)$$

sendo a Equação (3.12) para $\sigma = 0$ e a Equação (3.13) para $0 < \sigma \leq 8$. Os pesos gerados a partir de uma função Gaussiana são $w(m, n)$.

Os mapas de características são obtidas através das diferenças centro-vizinhança. O centro é um *pixel* da imagem em uma escala $c \in \{2, 3, 4\}$ e a sua vizinhança é um *pixel* correspondente em uma imagem em outra

escala $s = c + \delta$ com $\delta \in \{3, 4\}$. O processo de geração desses mapas são inspirados biologicamente em neurônios do córtex visual dos mamíferos, e as equações que definem matematicamente as diferenças centro-vizinhança são:

$$\mathcal{I}(c, v) = | I(c) \ominus I(v) | \quad (3.14)$$

$$\mathcal{RG}(c, v) = | (R(c) - G(c)) \ominus (G(s) - R(s)) | \quad (3.15)$$

$$\mathcal{BY}(c, v) = | (B(c) - Y(c)) \ominus (Y(s) - B(s)) | \quad (3.16)$$

$$\mathcal{O}(c, v, \theta) = | (O(c, \theta) \ominus O(s, \theta)) | \quad (3.17)$$

com $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$.

O modelo de [Benicasa \(2013\)](#) segue a hipótese de [Koch e Ullman \(1985\)](#), onde os mapas de cada característica alimentam o mapa de saliência. A região mais proeminente do mapa representa o ponto de maior saliência, independente se isso corresponde a um grande contraste de cor, intensidade ou orientação, pois nenhuma característica previamente conhecida é utilizada para guiar essa atenção, sendo portanto um modelo *bottom-up*.

Para formação desse mapa, os mapas de características são somados (\oplus), após terem sido normalizados ($\mathcal{N}(\cdot)$), gerando os mapas de intensidade (\bar{I}) e cor (\bar{C}). Essa normalização garante o descarte de alguns ruídos, ampliando o contraste das regiões salientes e inibindo regiões não contrastantes (como pode ser visto na [Figura 3.4](#)).

Os mapas de conspicuidades são formados como se segue:

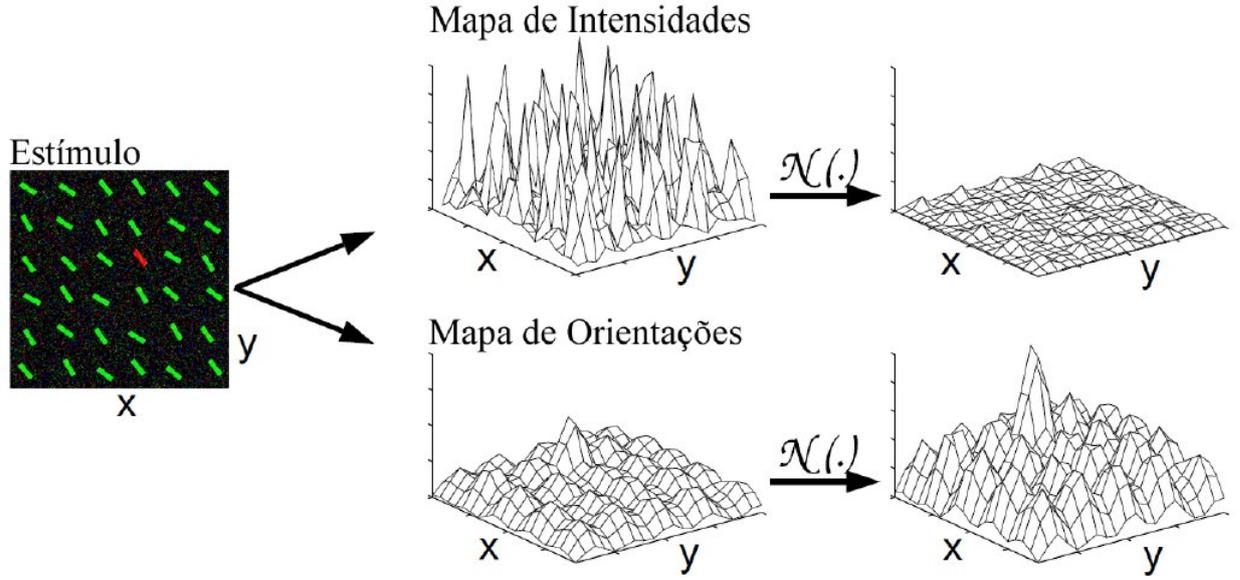


Figura 3.4: Normalização dos mapas de conspicuidades (Benicasa, 2013)

$$\bar{\mathcal{I}} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{I}(c, s)) \quad (3.18)$$

$$\bar{\mathcal{C}} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} [\mathcal{N}(\mathcal{RG}(c, s)) + \mathcal{N}(\mathcal{BY}(c, s))] \quad (3.19)$$

$$\bar{\mathcal{O}} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{O}(c, s, \theta)) \quad (3.20)$$

$$S = \frac{1}{3}(\mathcal{N}(\bar{\mathcal{I}}) + \mathcal{N}(\bar{\mathcal{C}}) + \mathcal{N}(\bar{\mathcal{O}})) \quad (3.21)$$

sendo S o mapa de saliência resultante de toda o cálculo.

3.4 Redes Neurais

3.4.1 *MultiLayer Perceptron (MLP)*

Organizações como a colônia de formigas, evolução natural das espécies, sistemas imunológicos, sistema nervoso, são exemplos de organizações que inspiram técnicas de computação. Algumas dessas técnicas, como as redes neurais artificiais e os algoritmos celulares têm sido, ao longo das últimas

décadas, aplicadas na resolução de problemas reais e têm se mostrado eficientes a problemas de difícil solução para a computação tradicional (de Padua Braga et al., 2000).

O modelo e o funcionamento do cérebro humano são a inspiração para as Redes Neurais Artificiais (RNAs). A principal semelhança entre as RNAs e o cérebro está no extenso procesamento paralelo apresentado e da independência de qualquer neurônio isolado. Assim, mesmo se um pequeno número de neurônios de uma rede forem danificados, ela continua a funcionar com pouca perda de desempenho. O neurônio, que representa a base do sistema nervoso (Kandel et al., 1997) é mostrado na Figura 3.5. O neurônio possui como função principal receber, processar e transmitir informações (Machado, 2000).

Com os conhecimentos da época, Warren McCulloch e Walter Pitts propuseram o primeiro modelo matemático de um neurônio. Esse Neurônio ficou restrito a representação computacional do neurônio biológico e não considerava o que ocorria no cérebro. Esse neurônio é composto por diversas entradas (os dendritos) que são ponderados pelos pesos (sinapses) e a saída (axônio) é obtida após um somatório. Na figura, θ representa o limiar de ativação do neurônio. Matematicamente, a saída do neurônio é dada por:

$$\text{sum}_{i=1}^n(x_i w_i) \geq \theta \quad (3.22)$$

$$v = \left(\sum_{i=1}^n x_i w_i \right) + \text{bias} \quad (3.23)$$

$$y = \varphi(v) \quad (3.24)$$

onde n representa o número de entradas (dendritos), x as entradas, w os pesos associados as sinápses (quando é w positivo a sinápsse é excitatória, quando negativo é inibitória) e θ representa o limiar de ativação. Matematicamente, a Equação (3.22) é o produto escalar (também conhecido como produto interno $x.w$) do vetor de pesos w com o vetor de entradas x .

Como o modelo de McCulloch e Pitts era muito limitado, modelos mais complexos, como o *perceptron* por exemplo, foram propostos. Atualmente, o modelo não-linear baseado no *perceptron* (Figura 3.6) é o mais usado pela

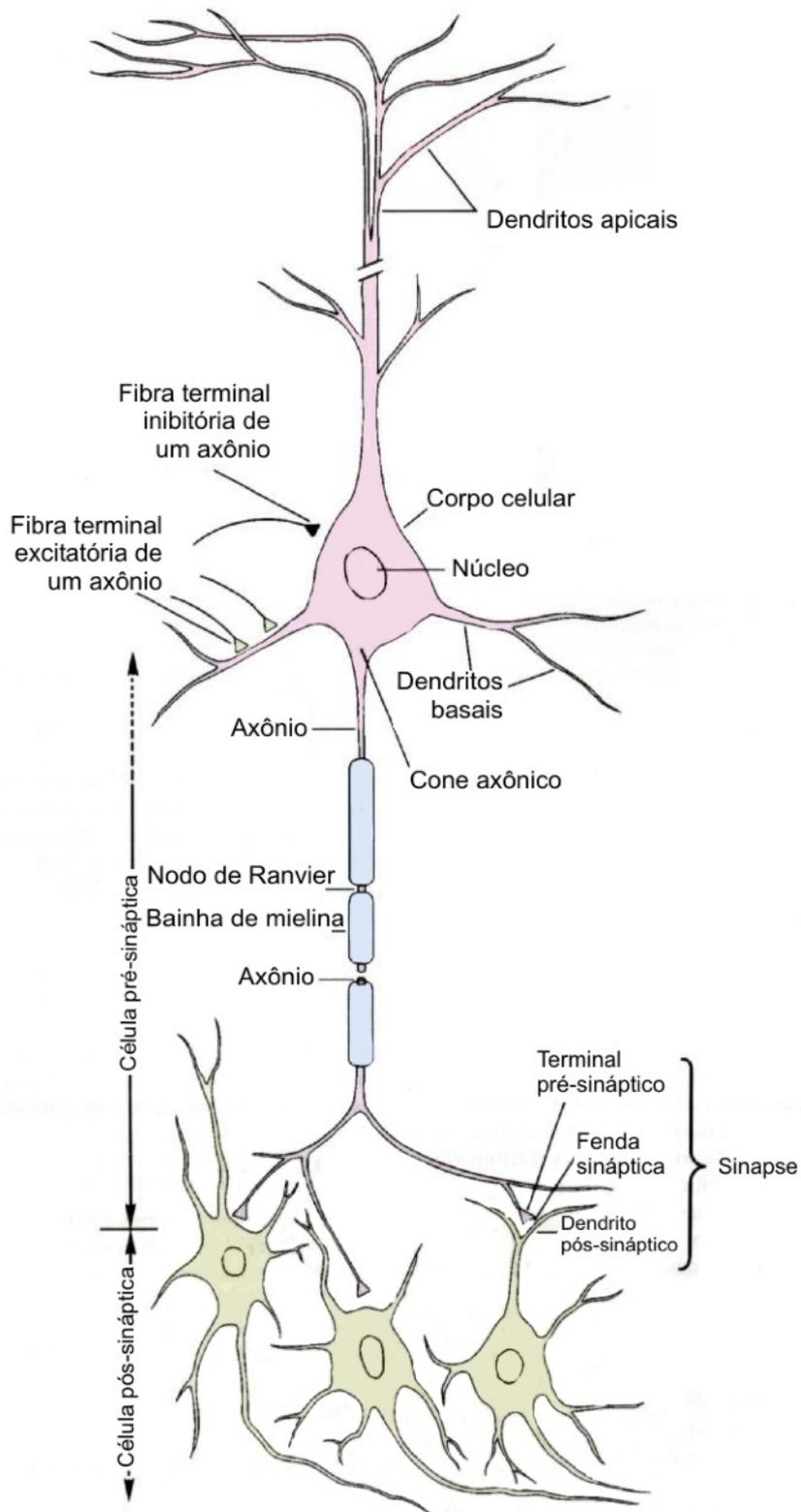


Figura 3.5: Morfologia do Neurônio Biológico (Kandel et al., 1997)

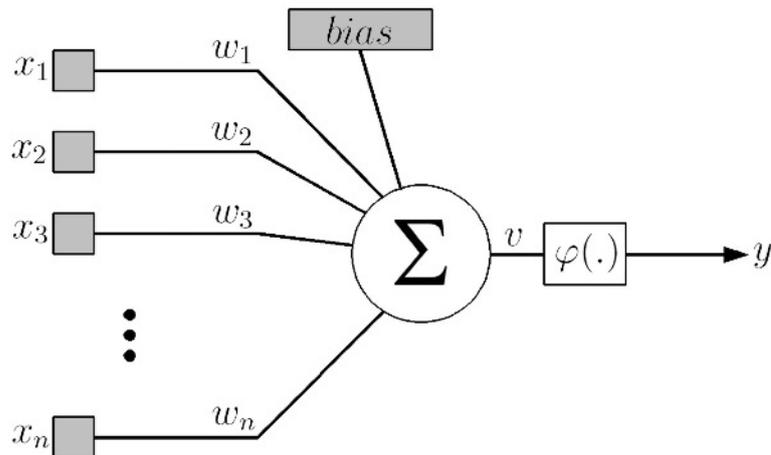


Figura 3.6: O Neurônio não-linear (Quiles, 2004)

comunidade de redes neurais. No neurônio não-linear, x_i representa a entrada presente na sinápsis i do neurônio, w_i o peso associado aquela sinápsis. v é o somatório das entradas ponderadas pelos pesos acrescido do termo *bias* (Equação 3.23) que é o limiar de ativação. $\varphi(\cdot)$ é a função de ativação do neurônio e y a sua saída (Equação 3.24).

Existem diversas funções de ativação possíveis, dentre as quais as mais utilizadas são a função sigmóide logística 3.25 e a tangente hiperbólica 3.26. A principal vantagem da utilização destas funções, segundo Haykin (2001) está na sua garantia da derivação, o que permite a construção de algoritmos que dependem dessa derivada para o cálculo do gradiente, como é o caso do algoritmo de retropropagação do erro (Seção 3.4.2).

$$\varphi(v) = (1 + \exp(-av))^{-1} \quad (3.25)$$

$$\varphi(v) = \tanh(v) \quad (3.26)$$

As redes neurais *perceptron* de múltiplas camadas (MLP) (Figura 3.7) são uma generalização do modelo de uma única camada. A arquitetura da MLP, apresentada na Figura 3.7, mostra a organização por camadas da rede. A primeira camada, chamada camada de entrada, simplesmente recebe os sinais e transmite para a próxima camada, não sendo constituída de neurônios. O sinal é propagado através da rede (pelas camadas $l=0,1,\dots,L$),

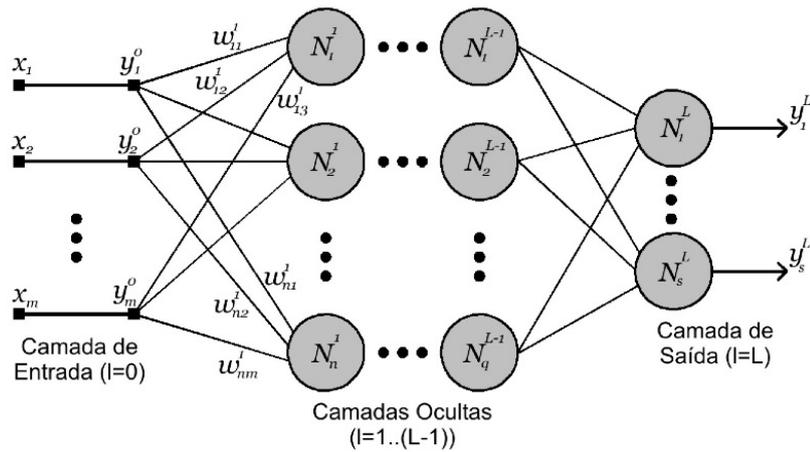


Figura 3.7: Arquitetura da Rede Neural MLP (Quiles, 2004)

sofrendo a ponderação do peso sináptico de cada camada (w_{ij}). A soma de todas as entradas (y_j) ponderadas é chamada de campo local induzido (v) e é representada pela Equação (3.27).

$$v_j^l = \left(\sum_{i=1}^m w_{ji}^l y_i^{l-1} \right) + bias_j^l \quad (3.27)$$

O sinal de saída do neurônio (N_i^l) é computado pela função de ativação $\varphi(\cdot)$ (Equação 3.28).

$$y_j^l = \varphi(v_j^l) \quad (3.28)$$

O problema principal com as MLPs era o ajuste de pesos nas camadas ocultas, o que dificultava a convergência da rede. Para resolver esse problema, os pesquisadores Rummelhart, Hintno e Willians propuseram em 1986 o algoritmo de retropropagação do erro (*backpropagation*).

3.4.2 BackPropagation

Um dos algoritmos mais importantes desenvolvidos para redes neurais, tanto por seu valor histórico quanto pelo seu poder computacional, é o *backpropagation*. Este algoritmo baseia-se na regra de aprendizado proposta por Widrow (chamada regra delta) e utiliza o gradiente descendente da superfície do erro, com a função de minimizar o erro quadrático.

O erro é retropropagado, ou seja, o erro da camada atual é uma estimativa dos erros das camadas anteriores. Por esse motivo, o numero excessivo de camadas ocultas deve ser evitado, pois a precisão do erro é diminuída (de Padua Braga et al., 2000). O algoritmo é composto por duas fases: A fase *feedforward*, onde o sinal de entrada é propagado para frente, e a fase *backward*, onde o erro é propagado para trás.

O erro é obtido fazendo-se a comparação da resposta da rede com a resposta esperada (ou saída desejada), no tipo de aprendizado chamado Aprendizado Supervisionado. Esse tipo de aprendizado é recomendado quando o trabalho é facilmente rotulável, como no caso da definição de cores.

O algoritmo 1 apresenta o *backpropagation*. A Equação (3.29) apresenta a entrada da rede. A saída dos neurônios da camada $l = 1$ é gerada pelas Equações (3.27) e (3.28) respectivamente. Camada a camada esse procedimento é executado até que o sinal atinja a camada de saída $l = L$ gerando o vetor $Y^L(p)$. Esse vetor é comparado ao vetor contendo os valores de saída desejados $D(p)$ para o padrão p apreendido (Equação 3.30) obtendo assim, o erro e_j para cada neurônio j da camada de saída L .

O erro quadrático total apresentado pela rede para o padrão de entrada p é definido pela por $E(p)$ (Equação 3.31) e o erro médio quadrático para todos os padrões é definido pela Equação (3.32). Os pesos w da rede são ajustados pela Equação (3.31). Para tal, camada a camada, o erro é retropropagado calculando-se os gradientes δ para cada neurônio (Equação 3.33). Os gradientes para os neurônios das camadas ocultas são calculados pela Equação (3.34). Após o calculo do gradiente, o ajuste dos pesos é feito pela Equação (3.35) onde η representa o coeficiente de aprendizagem. Vale ressaltar que um valor muito grande do η não garante a convergência do sistema.

$$y_j^0 = x_j \quad (3.29)$$

$$e_j(p) = d_j(p) - y_j^L(p) \quad (3.30)$$

$$E(p) = \frac{1}{2} \sum_{j \in L} e_j^2(p) \quad (3.31)$$

$$E_{med} = \frac{1}{P} \sum_{p=1}^P E(p) \quad (3.32)$$

$$\delta_j^l = e_j^L \varphi_j'(v_j^L) \quad (3.33)$$

$$\delta_j^l = \varphi_j'(v_j^l) \sum_k \delta_k^{l+1} w_{kj}^{l+1} \quad (3.34)$$

$$w_{ji}^l(t+1) = w_{ji}^l(t) + \eta \delta_j^l y_i^{l-1} \quad (3.35)$$

Algoritmo 1: Algoritmo de Retropropagação Padrão (Quiles, 2004)

para todo i e j das camadas $l = 1..L$ **faça**

$w_{ij}^l = \text{valor-aleatório}()$

fim para

enquanto não convergir **faça**

para todo Padrão $X(n)$ **faça**

Apresente $X(n)$ a RNA e obtenha a saída $Y(n)$

Calcule o erro $e(n)$ (Equação 3.30)

para todo Neurônio j da camada $l = L$ até a camada $l = 1$ **faça**

Calcule o Gradiente δ_j^l da seguinte maneira:

se Neurônio $j \in L$ **então**

$$\delta_j^l = e_j^L \varphi_j'(v_j^L)$$

senão

$$\delta_j^l = \varphi_j'(v_j^l) \sum_k \delta_k^{l+1} w_{kj}^{l+1}$$

fim se

$$\Delta w_{ji}(t) = \eta \delta_j^l y_i^{l-1}$$

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t)$$

fim para

fim para

fim enquanto

3.4.3 Rede LEGION

A Rede LEGION (Rede Oscilatória Localmente Excitatórias Globalmente Inibitória *Locally Excitatory Globally Inhibitory Oscillator Network* (Wang, 1995) é uma arquitetura muito utilizada nos últimos anos, principalmente pela sua alta velocidade de sincronismo e dessincronismo dos grupos de osciladores.

A arquitetura básica da rede é composta por três componentes principais: um inibidor global, acoplamentos excitatórios locais e osciladores neurais. Os objetos da cena são representados pelos grupos de osciladores, que são sincronizados pelos acoplamentos excitatórios. Já o inibidor gera a

dessincronização dos grupos distintos, que não representam o mesmo objeto na cena, gerando um mecanismo de cooperação local e inibição global. Em sua proposição inicial (Wang, 1995), o modelo é formado por uma rede de osciladores de relaxamento, sendo compostos pela variável excitatória x_i e pela variável inibitória y_i definidas como se segue:

$$\dot{x}_i = 3x_i - x_i^3 + 2 - y_i + \mathcal{I}_i + S_i + \rho \quad (3.36)$$

$$\dot{y}_i = \epsilon(\alpha(1 + \tanh(x_i/\beta)) - y_i) \quad (3.37)$$

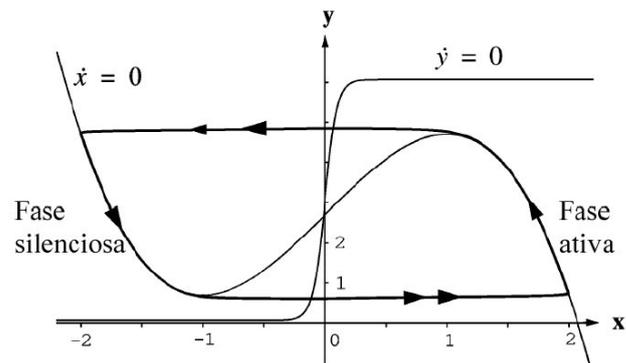
sendo I_i o estímulo externo ao oscilador i , S_i os acoplamentos, α, β e ϵ parâmetros do modelo (ϵ é normalmente uma constante positiva de valor pequeno), e ρ um sinal de ruído. Um oscilador de relaxamento típico quando I_i for constante e os termos S_i e ρ são eliminados.

As isóclinas nulas (quando $\dot{x}_i = \dot{y}_i = 0$) representam, respectivamente, uma função cúbica e uma função sigmóide. Quando $\mathcal{I}_i > 0$ o modelo representa um oscilador de tempo limite, oscilando em duas fases de valores de x . um mais elevado e a segunda de valores baixos. Essas fases são denominadas ativa e silenciosa e o oscilador estará no modo *disparando*. A transição entre as fases acontece muito rapidamente em uma fase denominada *jumping*. Quando não há estímulo ($\mathcal{I}_i < 0$) o sistema está num ponto de equilíbrio estável, podendo ser induzido a oscilar por meio de estímulos recebidos pelos acoplamentos. Neste caso o oscilador é considerado *excitável*. A figura 3.8 demonstra esses dois momentos do oscilador.

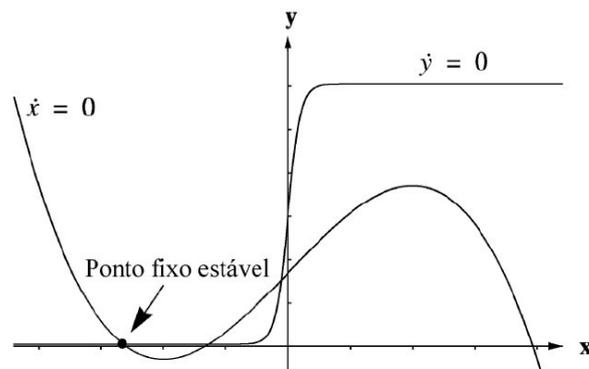
Considerando w_{ik} como a força de acoplamento entre os osciladores i e k , θ_x o limiar de ativação dos vizinhos por um oscilador, w_z acoplamento do sistema (S_i) a força da ligação entre o oscilador e o inibidor z e N_i a vizinhança de interação do oscilador, o acoplamento do sistema é dado por:

$$S_i = \sum_{k \in N_i} w_{ik} H(x_k - \theta_x) - w_z H(z - \theta_z) \quad (3.38)$$

A função $H(x)$ é a função de Heaviside, onde $H(x) = 1$ se $x \geq 0$ e $H(x) = 0$ caso contrário. Essa função, utilizada em Benicasa (2013), é uma evolução do modelo original, possibilitando a distinção de regiões maiores (prováveis objetos) e regiões menores (provável ruído). A dinâmica do inibidor global é



(a)



(b)

Figura 3.8: Modos do Oscilador de relaxamento: (a) Modo ativo e (b) modo excitável (Wang, 1995)

definida por:

$$\dot{z} = \phi(\sigma_\infty - z) \quad (3.39)$$

onde o parâmetro ϕ controla a velocidade de reação do inibidor. $\sigma \equiv 0$ caso nenhum oscilador tiver ultrapassado o limite θ_x e $\sigma \equiv 1$ se pelo menos um oscilador tiver ultrapassado. Nesse caso, z é estimulado e se aproxima de 1, passando a atuar como inibidor da rede quando z superar θ_z .

3.4.4 Mapa SOM

O objetivo principal do Mapa Auto-Organizado (*Self-Organizing Maps* - SOM) é organizar dimensionalmente os sinais de entrada em grupos (*clusters*) (Haykin, 2001). Esse mapa é responsável por transformar os dados de entrada (características primitivas e respostas da rede MLP) em um conjunto de dados bidimensional.

De acordo com Kohonen (2001), dado um vetor de peso $m_i = [\mu_{i1}, \mu_{i2}, \dots, \mu_{in}] \in \mathfrak{R}^n$ onde \mathfrak{R}^n representa o conjunto de entrada, μ_{in} representa a dimensão n do neurônio i e m_i tem seus valores inicialmente aleatórios. O vetor de entrada $p = [\xi_{i1}, \xi_{i2}, \dots, \xi_{in}] \in \mathfrak{R}^n$, para que aconteça competição entre os neurônios, é conectado paralelamente a todos os neurônios do mapa SOM através de m_i . Assim, o vetor de pesos que tiver o valor mais similar ao de entrada será o vencedor. Em Benicasa (2013) é sugerido o uso da menor distância Euclidiana ($\|p - m_i\|$), dada por:

$$d(p, m_i) = \|p - m_i\| = \sqrt{\sum_{j=1}^n (p_j - m_{ij})^2} \quad (3.40)$$

considerando n o número de dimensões. Essa distância está diretamente relacionada com o neurônio vencedor, que é definido por:

$$c = \arg \min \|p - m_i\| \quad (3.41)$$

A auto-organização da rede ocorre devido ao relaxamento local, que é, no período de aprendizado da rede, a atualização dos pesos do neurônio vencedor e de seus vizinhos próximos. Sendo t um numero inteiro que representa uma coordenada discreta de tempo, $\alpha_k(t)$ a taxa de aprendizado, e $h_{ci}(t)$ a função de relaxamento, essa atualização é dada por:

$$m_i(t+1) = m_i(t) + \alpha_k(t)h_{ci}(t)[p(t) - m_i(t)] \quad (3.42)$$

Para que o sistema possa convergir, $h_{ci} \rightarrow 0$ e $\alpha_k \rightarrow 0$ quando $t \rightarrow \infty$ (Kohonen, 2001). A função de relaxamento é definida, no trabalho de

Benicasa (2013), nos termos de uma função Gaussiana:

$$h_{ci}(t) = \exp\left(\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right) \quad (3.43)$$

onde $r_c \in \mathbb{R}^2$ e $r_i \in \mathbb{R}^2$ representam as posições dos neurônios c e i , e $\sigma(t)$ define a largura da região de vizinhança (raio). Segundo Kohonen (2001), normalmente $\sigma \leftarrow 0$ quando $t \leftarrow \infty$. A precisão da função de tempo é dada por:

$$\alpha_k(t+1) = \alpha_k(t)0.9\left(1 - \left(\frac{t}{n_{it}}\right)\right) \quad (3.44)$$

$$\sigma_k(t+1) = \sigma_k(t)0.9\left(1 - \left(\frac{t}{n_{it}}\right)\right) \quad (3.45)$$

onde n_{it} é o número de iterações para decrementar α e σ até zero.

3.5 Considerações finais

Neste capítulo foram demonstrados os métodos escolhidos para a implementação desta dissertação. Para o reconhecimento de objetos foi usada a técnica de Atenção Visual, por meio da combinação de mapa SOM com as Redes Neurais LEGION e MLP. No próximo Capítulo, será demonstrada a metodologia proposta e a implementação no robô humanoide NAO.

Metodologia

Neste capítulo serão demonstrados como os métodos, descritos no capítulo anterior, foram utilizados e integrados aos sistemas e bibliotecas da Aldebaran, que são utilizados pelo robô NAO. O robô humanoide, NAO, bem como suas funcionalidades são também apresentadas. Além disso, a arquitetura utilizada para o modelo de atenção visual é apresentada juntamente com o classificador utilizado para rotular as classes encontradas pela rede.

4.1 Robô NAO

4.1.1 Hardware

O robô NAO é um humanoide de cerca de 60 cm de altura, com massa em torno de 5 Kg, assemelha-se ao tamanho de uma criança e possui 25 graus de liberdade cujos elementos chave são os motores elétricos e atuadores; rede de sensores, incluindo duas câmeras, quatro microfones, sonar telemétrico, dois emissores e receptores IR; uma placa de inércia, nove sensores táteis e oito sensores de pressão; vários dispositivos de comunicação, incluindo o sintetizador de voz, luzes LED, e dois alto-falantes de alta fidelidade; ATOM 1,6 GHz Intel (localizado na cabeça) que executa um kernel do Linux e suporta middleware da Aldebaran (Naoqi), em segundo lugar CPU (localizada no tronco); bateria 27,6 W-h que fornece ao NAO em torno de 1,5 horas de

autonomia, dependendo do uso. Desenvolvido por uma empresa francesa, Aldebaran Robotics, este robô passou a possuir maior expressão junto à comunidade robótica por ter sido eleito para substituir o extinto robô Aibo da Sony Corporation em uma categoria da competição mundial de futebol de robôs, RoboCup Standard Platform League .

O robô tem sido utilizado em mais de 550 universidades e laboratórios de pesquisa espalhados pelo mundo. Robótica, Inteligência Artificial, Ciência de Computação divergindo para Sociologia e Cuidados Médicos são algumas das áreas de pesquisa onde esse robô tem sido utilizado.

Esta arquitetura permite a interação com outros robôs ou seres humanos por meio de distintas formas, as quais podem ocorrer pela fala, localização sonora, padrão visual e detecção de corpo, detecção de faces, detecção de obstáculos e efeitos visuais em LEDs que possui. O robô possui um sistema multimídia com dois alto-falantes (localizados nas laterais da cabeça), quatro microfones, duas câmeras digitais (dois orifícios, um na "testa", outro na "boca"), sensores táteis na cabeça e mãos, *bumpers* nos pés, sonares no seu torço, etc. A Figura 4.1 é mostrado a versão mais atual do robô NAO.

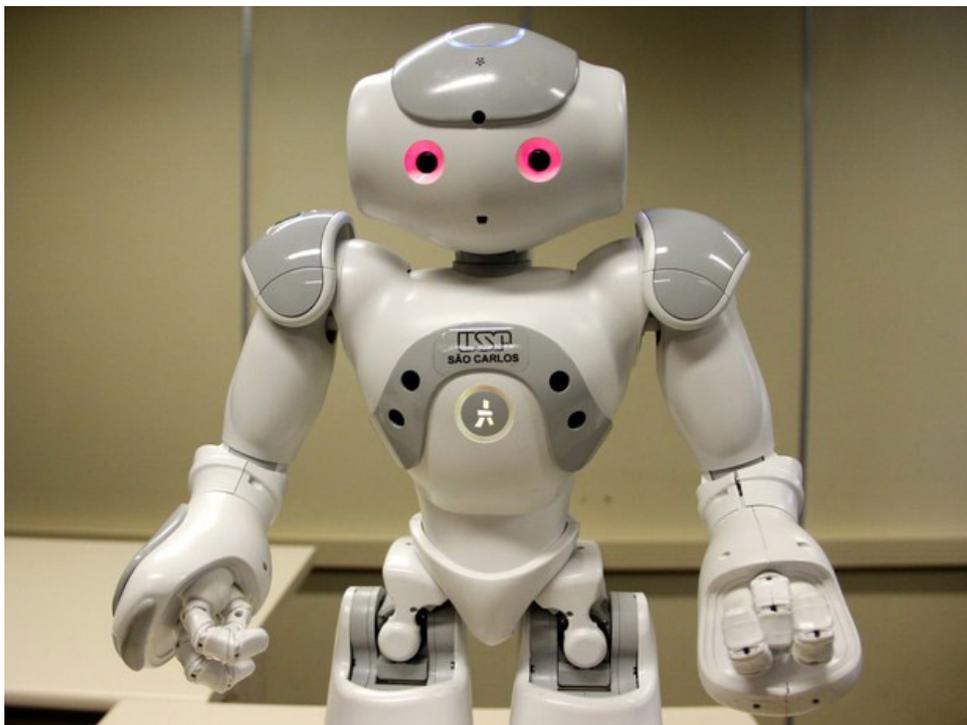


Figura 4.1: Robô NAO do laboratório LAR-USP

Em sua cabeça, na qual está instalada a CPU, o robô NAO tem um sistema operacional próprio, que permite a execução de comportamentos autônomos.

Esse sistema é chamado OpenNAO, baseado na distribuição de software livre GNU/Linux do Gentoo, e foi especificamente desenvolvido para esse robô. O OpenNAO provê diversas bibliotecas (proprietárias e livres), mas o software principal do robô é o NAOqi, pois é através dele que o robô se move, acende os LEDs e demais funções que são importantes para a interação que é esperada neste trabalho de Mestrado.

4.1.2 NAOqi

Juntamente com o robô, recebe-se alguns softwares padrão para criação de novos comportamentos e controle de todas as funções do robô, a partir de um computador. Esses softwares são o *Monitor* e o *Choreograph*. Enquanto o primeiro apenas permite algum *feedback* das câmeras do robô, o segundo, importante para esta pesquisa, é uma linguagem de programação baseada em blocos. Com o *Choreograph* é possível criar e testar animações e comportamentos de forma muito simples, testando primeiro em um robô simulado para depois ser utilizado no robô real. Esses comportamentos criados são escritos em uma linguagem gráfica passível de ser interpretada e executada pelo NAOqi sem a necessidade de se conhecer uma única linha de código. Portanto, toda a interação Homem-Robô foi feita dentro do *Choreograph*, como as falas, os movimentos de braços, pernas e cabeça, a coloração dos LEDs (vermelha para o caso de erro e azul intermitente para o caso de acerto). Por meio de bibliotecas padrão proprietárias da Aldebaran, todos esses comportamentos salvos no robô podem ser acessados diretamente por programas escritos em C++.

4.1.3 C++ SDK

Toda a implementação da atenção visual fora feita em C++, e era necessário integrá-la ao robô NAO. Para isso, existe uma SDK (*Software Development Kit* - Kit de desenvolvimento de Software) para ser utilizada com o sistema do NAOqi. A partir desta SDK, um novo módulo pode ser criado e salvo no robô. Essa SDK é chamada NAOqi API, disponível para computadores em 8 diferentes linguagens, porém apenas C++ e Python são interpretadas pelo robô. A *framework* C++ é a mais completa e a única que permite a escrita de código em *real-time*, o que é importante para nossa abordagem.

Para compilar os códigos para o robô, é recomendado pela própria Aldebaran a utilização do CMake (um sistema multiplataforma para realizar

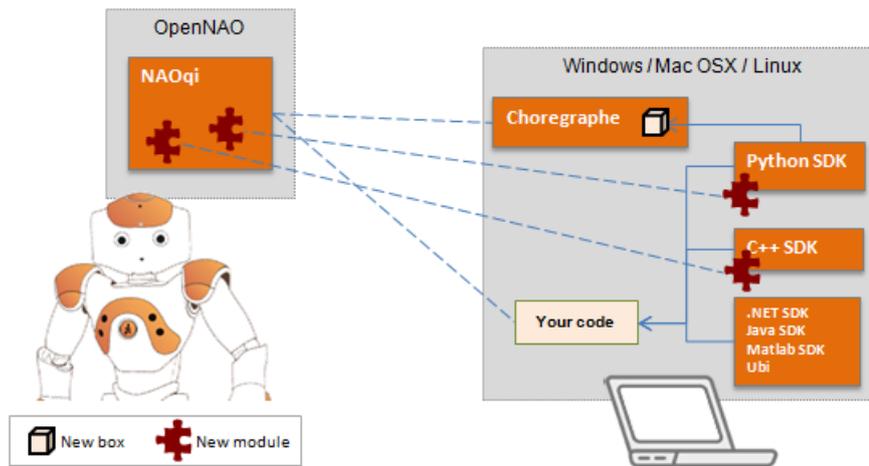


Figura 4.2: Como o Software do NAO funciona [Robotics \(2014\)](#)

geração automatizada) juntamente com a *framework* qiBuild. O qiBuild gerencia as dependências entre os projetos e suporta o *cross-compilation*, permitindo a boa interpretação do código pelo OpenNAO.

É obrigatório, para a versão do robô NAO utilizado, a compilação utilizando arquitetura de 32-bits. Apesar disso, o código resultante irá funcionar corretamente tanto qualquer ambiente, seja 32 ou 64 bits. Também considerando a versão utilizada do robô, a *framework* C++ SDK suporta apenas a versão 2.3.1 da biblioteca OpenCV, uma das maiores e mais utilizadas bibliotecas de visão computacional *open source*, responsável tratar as imagens das câmeras do robô.

A Aldebaran recomenda a utilização da IDE (*Integrated Development Environment* - Ambiente de Programação) do QtCreator, pois algumas funções do QT estão inseridas no robô, tornando mais fácil a implementação de um novo módulo.

4.2 Arquitetura

Essa dissertação propõe uma estratégia de reconhecimento de figuras geométricas por meio de um mecanismo de atenção visual *bottom-up* com enviesamento *top-down*. Para tal é utilizado um Mapa de Atributo-Saliência (MAS) gerado a partir da retirada de características primitivas da imagem, de um mapa SOM e de informações prévias de objetos conhecidos das imagem, obtido através de uma rede MLP ([Benicasa, 2013](#)). O sistema é composto por três módulos: A verificação de regiões salientes, a atenção visual e a classificação. Essa arquitetura foi implementado para ser utilizado no robô

NAO. A Figura 4.3 sumariza essa arquitetura proposta.

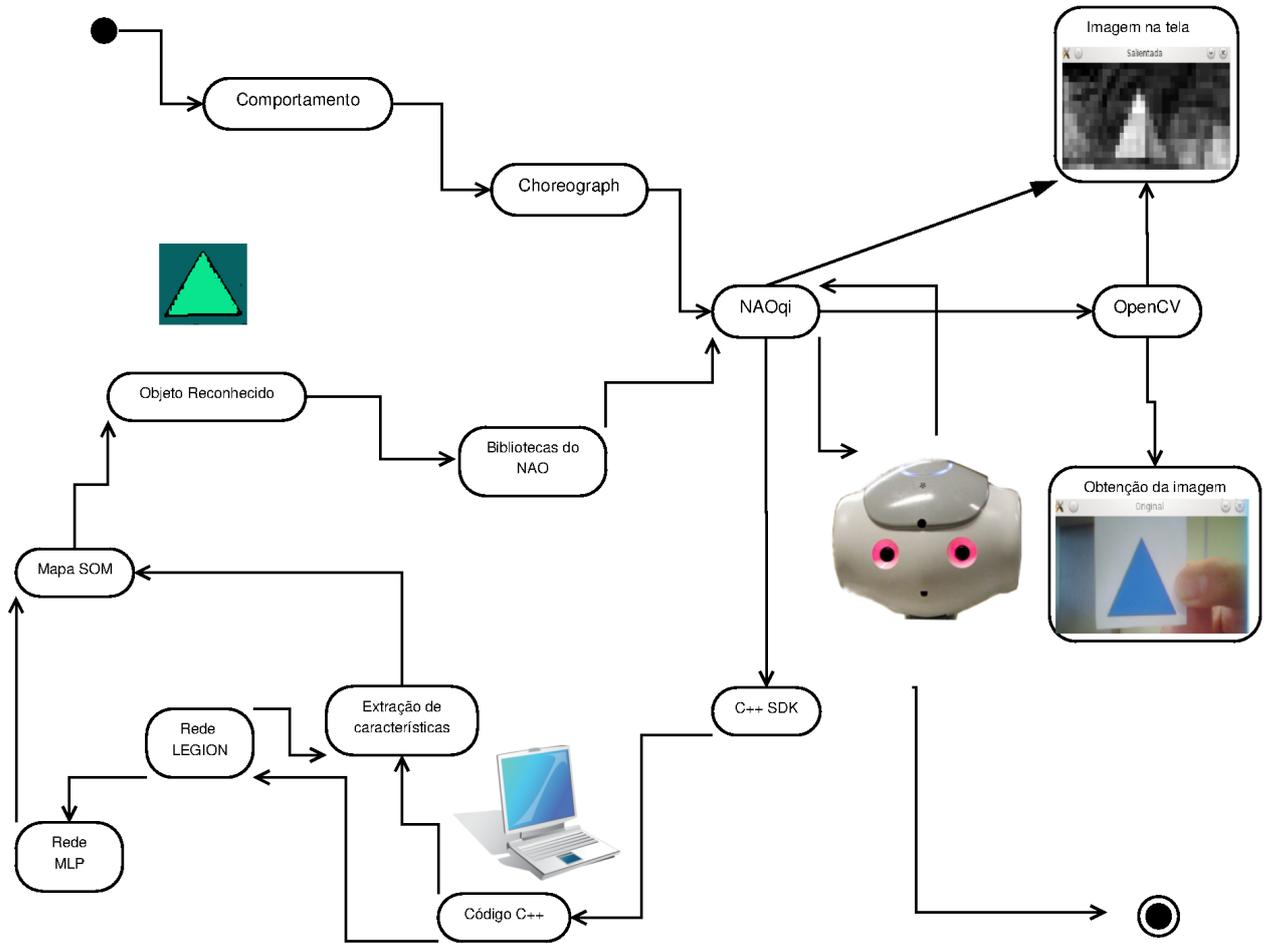


Figura 4.3: Fluxograma do Sistema Proposto

4.2.1 Descrição do Modelo

A segmentação para cenas reais, de acordo com as limitações de outros modelos descrita por Benicasa et al. (2013) e discutidas do capítulo 3, é feita através da rede LEGION.

O fluxo de informações funciona da seguinte maneira: paralelamente uma cena é apresentada para o módulo de extração de características primitivas (cor, orientação e intensidade) e para a rede LEGION. A saída da rede LEGION alimenta a rede MLP e o mapa SOM, gerando o reconhecimento do objeto e a geração do mapa de saliência.

Como descrito anteriormente, esta rede trabalha com modulação *bottom-up* e *top-down* para a atenção visual. Essa extração de características primitivas definem a informação *bottom-up* deste modelo, sendo consideradas a partir da

imagem de entrada de todo o sistema. A modulação *top-down* é possível por meio da rede MLP, e utiliza as informações sobre esses objetos previamente conhecidos para guiar a seleção da área mais saliente.

4.2.2 Modelo *bottom-up*

A extração das características primitivas ocorre a partir da imagem de entrada do sistema obtida pela câmera do robô, retirando-se informações dos canais r, g e b e as cores oponentes RG e BY , intensidades I e orientações, conforme descritos no capítulo 3.

Para a segmentação, a imagem inicial precisa ser dividida em um conjunto de segmentos, que representam objetos da cena e também alguns ruídos. Esse processo já diminui o tamanho da imagem a ser considerada, além de possibilitar que esses segmentos sejam passados para a rede MLP, para o reconhecimento do objeto e enviesamento *top-down* (Explicado na próxima seção). Essa segmentação é feita pela rede LEGION.

O modelo LEGION modificado, proposto em Benicasa (2013), é recomendável para utilização em cenas reais, pois atualiza o modelo de Wang (1995), introduzindo um termo de potencial lateral. A partir desse termo é possível diferenciar alguns ruídos de segmentos maiores que representam objetos. Esse mecanismo funciona atribuindo ao oscilador que está no centro de um segmento, ou seja, uma região homogênea, o *status* de "líder" do segmento. Caso o oscilador não represente uma região, mas sim *pixels* isolados, não terá esse *status*. Um "líder" está apto a receber alto potencial dos seus vizinhos, assim o processo de segmentação é conduzido de forma que só os líderes oscilem, inibindo as regiões isoladas (ruídos). A Figura 4.4 mostra exemplos de segmentos em uma imagem de um círculo reconhecido pelo NAO.

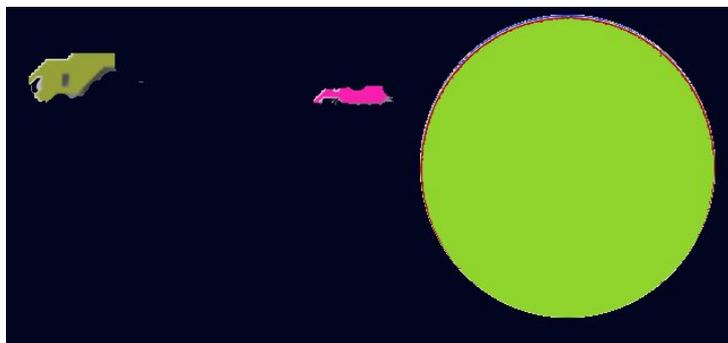


Figura 4.4: Segmentação do LEGION feita pelo NAO

4.2.3 Modelo *top-down*

A rede MLP classifica objetos previamente conhecidos da cena, e guia a seleção, inibindo segmentos que não representam objetos geométricos, alvo desta pesquisa. Para uma correta classificação dos objetos salientes apresentados em uma cena, a rede precisa ser treinada com um conjunto de exemplos que representa muito bem os alvos desejados.

Esse conjunto de treinamento é criado por meio de imagens extraídas a partir da cena. Os objetos desejados são apresentados a segmentação da rede LEGION, que devolve os segmentos encontrados. Todos esses segmentos, caso representem algum objeto, são classificados manualmente e apresentados a rede MLP, para que seja feito o treinamento e a definição dos pesos nas camadas da rede. A Figura 4.5 apresenta alguns objetos que podem ser reconhecidos pela MLP.

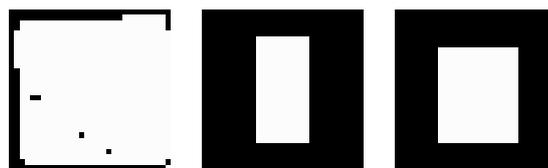


Figura 4.5: Exemplos do treinamento

Uma vez finalizado o treinamento, a rede está preparada para reconhecer novos objetos (segmentos vindos da rede LEGION) nunca antes vistos. O valor de saída da rede indica se o objeto foi rotulado como um daqueles apresentados anteriormente no período de treinamento ou não. Uma vez que tenha reconhecido um objeto, a saída encontrada serve para configurar o atributo referente ao valor de reconhecimento. Inicialmente esse valor é atribuído zero para todos os neurônios, e todos que estão relacionados a algum objeto que deve receber atenção recebe o valor 1, possibilitando a modulação baseada em objetos.

Já que esses os segmentos e os dados de treinamento possam ser muito diferentes, criando muitas rotulações incorretas, a pirâmide gaussiana foi utilizada inclusive para a criação dos conjuntos de treinamento. A partir dela, todos os objetos, independente de estarem mais próximos ou mais afastados da câmera (tornando os segmentos maiores ou menores para serem reconhecidos), ficam muito próximos de um mesmo tamanho, quando

retirados do oitavo e ultimo nível da pirâmide. Isso aumentou a acurácia do sistema, possibilitando um reconhecimento mais confiável dos objetos para servir de guia ao processo de seleção.

O Mapa SOM utiliza todos os dados primitivos extraídos da cena e a saída da MLP para seu treinamento. Os detalhes da implementação do mapa estão descritos no Capítulo 3.

4.3 Implementação no robô NAO

Seguindo a metodologia definida na Seção 4.1, toda a implementação foi feita em C++, utilizando a OpenCV. Para tal, a imagem obtida pela câmera do robô NAO foi salva em uma matriz do tipo Mat (tipo de dados da OpenCV). Essa matriz era convertida em uma outra matrix de Neurônios, onde as informações primitivas da imagem (canais de cores) eram salvas em posições separadas do neurônio, assim como informações do oscilador (se deveria ou não pulsar). Depois eram retiradas as informações de orientação, necessárias para o Mapa SOM, além das diferenças de cores *RG* e *BY*.

Com as características primitivas da imagem retiradas, inicia-se a criação dos oito níveis da pirâmide gaussiana, sempre dividindo a imagem pela metade para o próximo nível da pirâmide. O ultimo nível possível da Pirâmide Gaussiana, uma vez que não seria possível diminuir oito vezes uma imagem muito pequena, é enviado para a rede LEGION, para que seja feita a extração dos segmentos.

A implementação da rede LEGION aplicada a imagens reais requer um grande esforço computacional, o que atrapalha na obtenção dos segmentos (e devida rotulação) em tempo real. Assim seguimos o algoritmo proposto em Benicasa (2013), que tem como principal característica a existência de apenas 2 parâmetros a serem utilizados: W_z que é o peso do inibidor global e θ_p um limiar responsável pela formação dos osciladores líderes.

Com os testes realizados em Benicasa (2013) e corroborados nesta pesquisa, o aumento de W_z reflete na maior dificuldade para o agrupamento em uma região, o que guia o algoritmo para regiões mais homogêneas. Já a oscilação de θ_p influencia diretamente na formação dos líderes (quanto maior θ_p , maior dificuldade para formação) e, conseqüentemente, no número de regiões pulsantes (ou segmentos). Assim, foi concluído que a melhor abordagem eram altos valores de W_z e baixos valores de θ_p , o que gera um maior número de segmentos.

Com a segmentação concluída, esses segmentos são enviados para a rede MLP. A rede recebe a imagem (cada neurônio representando um pixel da imagem), e possui seis neurônios de saída, cada um representando uma figura geométrica diferente (triângulo, quadrado, retângulo, círculo e trapézio). Durante os testes, conforme descrito no Capítulo 5, essa metodologia precisou ser alterada para incluir outras figuras geométricas.

Após a rede reconhecer o objeto, e ter um valor de reconhecimento, esse valor foi incluído para a finalização do mapa SOM, que gera os mapas de saliência e é guiado para o foco de atenção e, conseqüentemente, para o objeto procurado. Após testes, foram obtidos os seguintes valores para os principais parâmetros de toda a rede:

- LEGION: $\theta_p = 400$ e $W_z = 5.0$
- SOM: $\alpha = 0.5$ e $\sigma = 1$
- MAS: $W_r = 2$ (O valor da intensidade foi descartado)

O fluxograma apresentado na Figura 4.6 resume todo o processo de atenção visual.

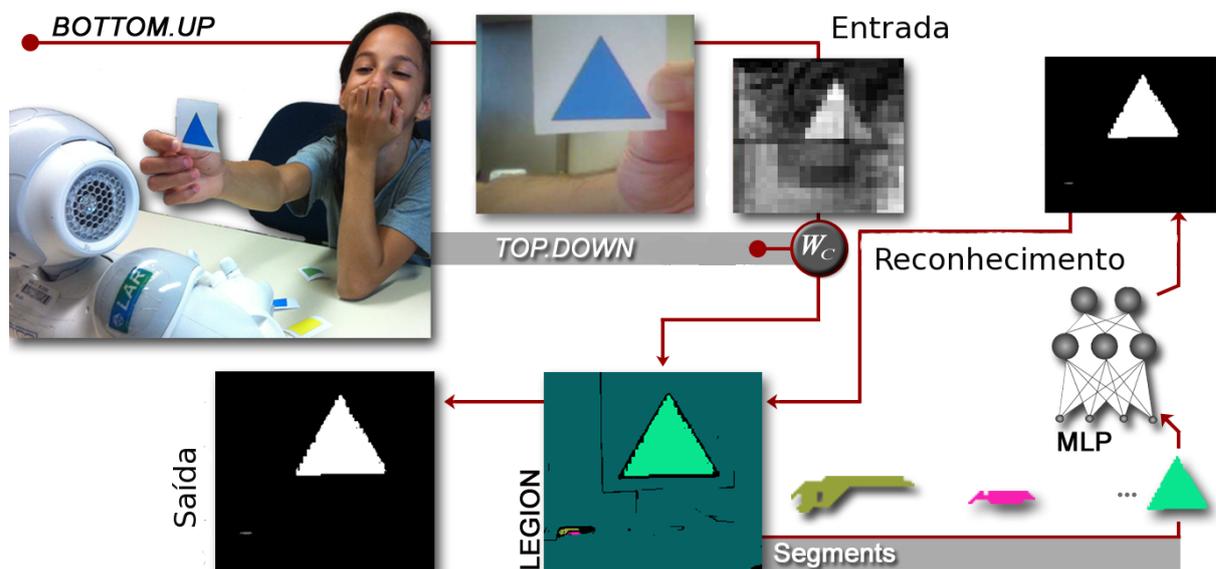


Figura 4.6: Fluxograma de funcionamento do sistema

4.3.1 Bibliotecas do NAO utilizadas

Com a atenção definida e o objeto reconhecido, o robô estava pronto para dar o *feedback* ao usuário final. Para que isso seja possível, algumas

bibliotecas da Aldebaran foram utilizadas no código final. Foram três bibliotecas de visão para obter as imagens e mostra-las na tela do computador: ALVideoDevice, ALImage e ALVisionDefinitions, uma biblioteca para gerenciamento de erros (ALError) e as duas bibliotecas mais importantes: a biblioteca para permitir as falas do robô (ALTextToSpeech) e a biblioteca para tratar os comportamentos pré-gravados no robô NAO (ALMotion).

O processamento embarcado no NAO, mesmo com todos os tratamentos utilizados nesta pesquisa, mostrou-se muito complexo para ser executado diretamente no robô. A resposta não seria obtida em tempo real e acarretaria em muitos momentos "vazios" entre as atividades do robô, o que poderia causar um desinteresse de uma criança, o que vai de encontro com os objetivos desta dissertação. Assim, optou-se para deixar o processamento sendo realizado apenas em um microcomputador, deixando o robô NAO apenas como atuador, isto é, exercendo tarefas que já estão disponíveis no mesmo, como fala, acendimento de luzes, entre outros.

4.4 Considerações Finais

Nesse capítulo, foi apresentada a metodologia utilizada para realização deste mestrado. O sistema foi construído utilizando os conceitos apresentados no capítulo anterior, unindo-se aos conceitos e bibliotecas do robô NAO. No próximo capítulo, serão apresentados os experimentos que foram construídos para serem realizados com o robô em ambiente escolar.

Experimentos

Neste capítulo são demonstrados os experimentos realizados visando verificar o desempenho do sistema proposto. O sistema de visão foi implementado com sucesso no robô NAO, e o robô foi levado à sala de aula como professor de alunos do ensino fundamental. Os diferentes cenários (atividades) para os testes são apresentados mostrando como o robô interagiu junto com as crianças. A avaliação dos experimentos foi feita pelos próprios alunos que participaram das atividades propostas. As respostas dos alunos e a observação durante os testes serão levados em consideração para a discussão dos resultados obtidos.

5.1 Cenário de testes

Para verificar a eficiência do sistema integrado proposto, alguns testes foram conduzidos com imagens sintéticas e outros em ambientes reais. Como descrito nos objetivos dessa dissertação (Capítulo 1) a ideia principal era utilizar o robô NAO dentro de sala de aula. Para tal algumas atividades foram propostas, com auxílio de professores da área da Educação e Ensino de Matemática. Foram conduzidos testes com 62 crianças, todas com idades entre 13 e 14 anos, vindas de Escolas Públicas, da cidade de São Carlos - SP. Os experimentos foram feitos em 3 etapas principais: 2 etapas de questionários, e uma etapa envolvendo a interação com o robô NAO. Para poder fazer uma comparação melhor elaborada, foi considerado ainda um

grupo de controle, no qual as crianças realizaram as mesmas atividades, porém sem a utilização do robô.

Durante os experimentos foi utilizada a métrica do *Mean Opinion Score* - (MOS) (Rec, 2000), para possibilitar uma avaliação subjetiva da experiência com o robô. Essa técnica oferece uma escala para medir a qualidade das interações com o NAO, além dos seus movimentos, fala e a habilidade de entender e reconhecer as figuras apresentadas. Essa escala é vai de 1 a 5, onde 1 significa que a interação foi muito ruim e 5 significa que ela foi excelente.

Os testes com as crianças ocorreram com a presença de um professor da escola, para que ele também pudesse expor sua visão da metodologia proposta. No primeiro momento, foi aplicado um questionário de "pré-atividade", com perguntas sobre figuras geométricas em geral. Esse questionário media apenas o conhecimento prévio do aluno sobre as figuras, para verificar se alguma das atividades com o robô estavam fora do escopo do aprendizado em sala de aula. Um vez finalizado o questionário, individualmente os alunos eram conduzidos para a interação com o robô NAO.

5.2 Testes com a presença do robô

Três diferentes cenários com o robô foram conduzidos. No primeiro e mais simples, a criança deveria mostrar uma imagem de um gato (Figura 5.1), feito com as peças do Tangram¹. Em todas as interações, o robô, graças ao software integrado desenvolvido, realizou todo o processo de atenção visual, para reconhecer todas as figuras geométricas presentes na imagem, e perguntou qual a quantidade de triângulos que a criança conseguia contar. Com as primeiras crianças, o robô aguardava que ela falasse a resposta, para utilizar o reconhecimento de voz que vem implementado no NAO, mas essa abordagem foi logo descartada por apresentar muitas falhas. O aluno, então, respondia via teclado no computador que estava conectado ao robô e que fazia o processamento da visão computacional, o que foi considerado inviável de ser feito diretamente no robô.

¹Tangram é um quebra-cabeças chinês formado por 7 peças, chamadas tan (5 triângulos, 1 quadrado e 1 paralelogramo), que juntas podem tomar diversas formas. O objetivo do quebra-cabeças é criar alguma forma, dada a silhueta de algum objeto ou animal, usando as 7 peças. É possível formar até 1700 figuras diferentes. O Tangram é usado na matemática como um facilitador no entendimento de figuras geométricas

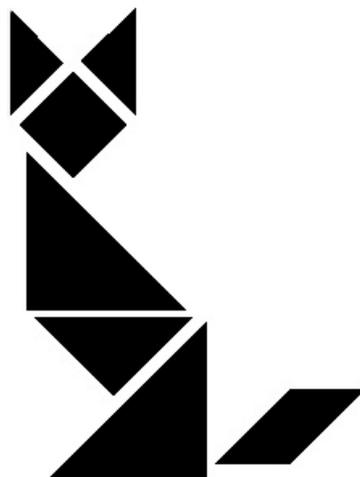


Figura 5.1: Gato feito com o Tangram

Em todas as atividades, o robô estava preparado para dar duas dicas no caso de uma resposta errada. No caso desta primeira atividade, as falas do robô eram:

- **Caso resposta errada na primeira tentativa:** "O triângulo é uma figura com três lados, você consegue encontra-los agora?"
- **Caso resposta errada na segunda tentativa:** "As orelhas do gato são formadas por triângulos. Tente mais uma vez."
- **Caso resposta certa em qualquer tentativa:** "Isso mesmo, parabéns! Agora escreva no papel quais outras figuras você consegue encontrar nessa imagem."

O caso da criança errar nas três tentativas não foi abordado nesse exemplo. Uma discussão do impacto dessa metodologia será feita na Seção 5.4.

Este primeiro experimento, juntamente com o questionário, são atividades mais simples, apenas para verificar o conhecimento da criança e as reações dos alunos com o robô. O nível de dificuldade foi incrementado em cada uma das atividades propostas. A Figura 5.2 apresenta um momento deste cenário.

No segundo experimento, foi proposta uma disputa entre os alunos e o robô, com um jogo de perguntas feitas pelo NAO. O robô perguntava por alguma figura específica, dando apenas algumas características, e esperava que a criança mostrasse essa figura para sua câmera. Caso a resposta fosse correta o robô abria os braços para o alto e piscava os LEDs do olho e do peito, em sinal de alegria pela resposta correta. Caso a criança mostrasse



Figura 5.2: Aluno mostra a figura do gato para o robô NAO

uma figura errada, o robô abaixava a cabeça e ficava com os LEDs vermelhos. Novamente, o robô estava preparado para dar duas dicas, no caso de respostas erradas. Nesta interação, a criança não precisava de utilizar o teclado, pois o robô não precisava de outra informação além da própria figura mostrada. Neste experimento, o robô perguntava por três diferentes figuras, e as suas falas eram as seguintes:

- **Primeira figura geométrica- Retângulo**

- **Pergunta inicial:** "Qual destas figuras geométricas na sua frente tem quatro lados?"
- **Caso resposta errada na primeira tentativa:** "A formula da área da figura que eu procuro é base vezes altura"
- **Caso resposta errada na segunda tentativa:** "A formula do perímetro dessa figura é duas vezes a soma de sua base vezes a altura".
- **Caso resposta certa em qualquer tentativa:** "Parabéns, você acertou! O retângulo é uma figura geométrica de quatro lados, onde sua área é a base vezes a altura e o perímetro é duas vezes a soma da base vezes a altura. Vamos para a próxima figura!"
- **Caso resposta errada na terceira tentativa:** "Que pena, a figura

que eu estava procurando era um retângulo. Mas não desanime, vamos tentar acertar a próxima figura!"

- **Segunda figura geométrica- Triângulo**

- **Pergunta inicial:** "Qual destas figuras geométricas na sua frente tem três lados?"
- **Caso resposta errada na primeira tentativa:** "Não não, essa não é a resposta certa. A formula da área da figura que eu procuro é base vezes altura dividido por dois"
- **Caso resposta errada na segunda tentativa:**"Hmm, ainda não. Essa figura ela pode ser classificada em três tipo: Equilátero, Escaleno e Isóceles".
- **Caso resposta certa em qualquer tentativa:** "Parabéns, você acertou! O triângulo é uma figura geométrica de três lados, onde sua área é a base vezes a altura dividido por dois e ele pode ser classificado em Isóceles, Escaleno ou Equilátero, de acordo com o tamanho dos seus lados. Vamos para a ultima figura!"
- **Caso resposta errada na terceira tentativa:** "Que pena, a figura que eu estava procurando era um triângulo. Mas vamos lá, vamos tentar acertar a ultima figura figura!"

- **Terceira figura geométrica- Quadrado**

- **Pergunta inicial:** "Preste atenção na pergunta. Qual destas figuras geométricas na sua frente tem quatro lados iguais?"
- **Caso resposta errada na primeira tentativa:** "Não não, essa não é a resposta certa. Veja, a área desta figura é o seu lado ao quadrado"
- **Caso resposta errada na segunda tentativa:**"Hmm, ainda não. Essa figura tem como perímetro quatro vezes o seu lado".
- **Caso resposta certa em qualquer tentativa:** "Parabéns, você acertou! O quadrado é uma figura geométrica regular de quatro lados, onde sua área é seu lado ao quadrado e seu perímetro é quatro vezes a medida do seu lado. Você mandou muito bem!"
- **Caso resposta errada na terceira tentativa:** "Que pena, a figura que eu estava procurando era um quadrado. Mas tudo bem, você mandou muito bem"



Figura 5.3: Jogo de perguntas com as crianças

Novamente o robô, que agia o tempo todo de forma autônoma, passava pelo processo de atenção e reconhecimento das figuras em todas as interações. Esse processo era feito repetidamente para comparação das diferentes posições em que a criança poderia mostrar as figuras geométricas para o robô. As Figuras 5.3 e 5.4 mostram diferentes crianças durante este experimento.

No terceiro e último cenário, as crianças puderam interagir com o robô por meio de um pequeno jogo, agora com as figuras geométricas sobrepostas. Novamente, foi necessário mostrar uma imagem para o robô (Figura 5.5(a)) que, por sua vez, realizou o processamento das imagens e pode fazer a contagem de quantas figuras diferentes foram encontradas na imagem. O robô perguntava quantas figuras o aluno conseguia identificar na imagem e aguardava a resposta, dada via teclado. A contagem de figuras foi possível graças aos inibidores de retorno da rede LEGION, que não permitiam que um mesmo segmento já reconhecido (ou descartado como ruído) fosse enviado duas vezes para o robô. Uma vez que um segmento era considerado um objeto reconhecido, o contador de figuras era incrementado, e o robô tinha um valor final da quantidade de figuras encontradas.

Nestas atividades, especificamente, o robô dava apenas uma outra chance para que fosse encontrado o número correto de figuras geométricas. Isso foi necessário para encurtar o tempo do experimento, pois a criança, uma vez



Figura 5.4: Com a resposta errada, o robô se agacha e fica vermelho

que se sentia desafiada pelo robô, acabava demorando muito para responder, buscando a resposta correta. Na segunda chance, o robô apenas avisava que existiam algumas figuras escondidas, mas não dava maiores dicas sobre como encontrá-las. A criança, acertando ou não na segunda chance, tinha um *feedback* imediato do robô, que mostrava onde estavam as figuras escondidas, como pode ser visto na Figura 5.5(b).

A falta de dicas e a pequena quantidade de chances foram propostas também para medir a reação com a segunda parte do experimento, que era basicamente a mesma aplicação, porém em uma imagem um pouco diferente (Figura 5.6). A mesma quantidade de chances e as mesmas ações do robô foram consideradas para esta segunda parte. As Figuras 5.7 e 5.8 apresentam crianças durante este segundo experimento.

5.3 Questionário final - Avaliação das atividades pelos alunos

Finalizando o conjunto de experimentos, foi apresentado para o aluno um segundo questionário, em uma outra sala, sem o robô. Neste questionário, algumas perguntas sobre as figuras geométricas e figuras "escondidas (sobrepostas)" foram apresentadas, para medir se o conceito trabalhado com

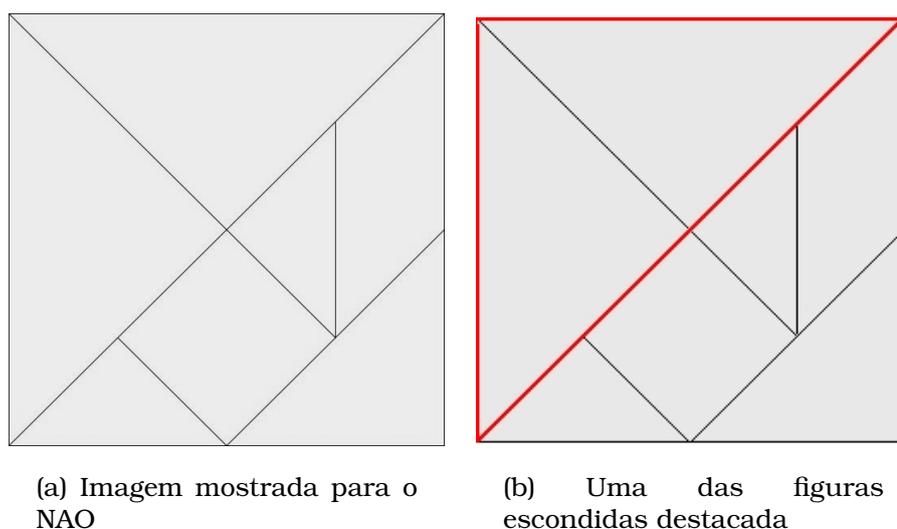


Figura 5.5: Exemplo da primeira imagem

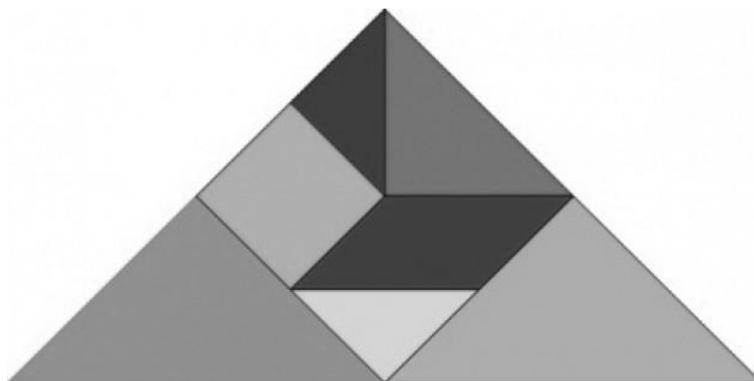


Figura 5.6: Segunda imagem, um pouco mais complexa

o robô tinha sido absorvido por parte das crianças. Foram consideradas perguntas sobre as formulas passadas pelo robô e também novas imagens sobre figuras sobrepostas, onde as crianças deveriam encontrar possíveis figuras escondidas dentre as que estavam no questionário.

Todas as interações com os alunos foi salva em imagens e vídeos, e todas as expressões e respostas das crianças foram consideradas para os resultados finais dessa pesquisa. Ao final do questionário, além de um espaço para qualquer observação sobre todo o processo dos experimentos (essa parte pôde ser respondida inclusive pelos professores que acompanhavam os alunos durante os experimentos), foram propostas algumas perguntas para avaliação do MOS, que são:

- 1) O que você achou do tempo de resposta e ação do robô?

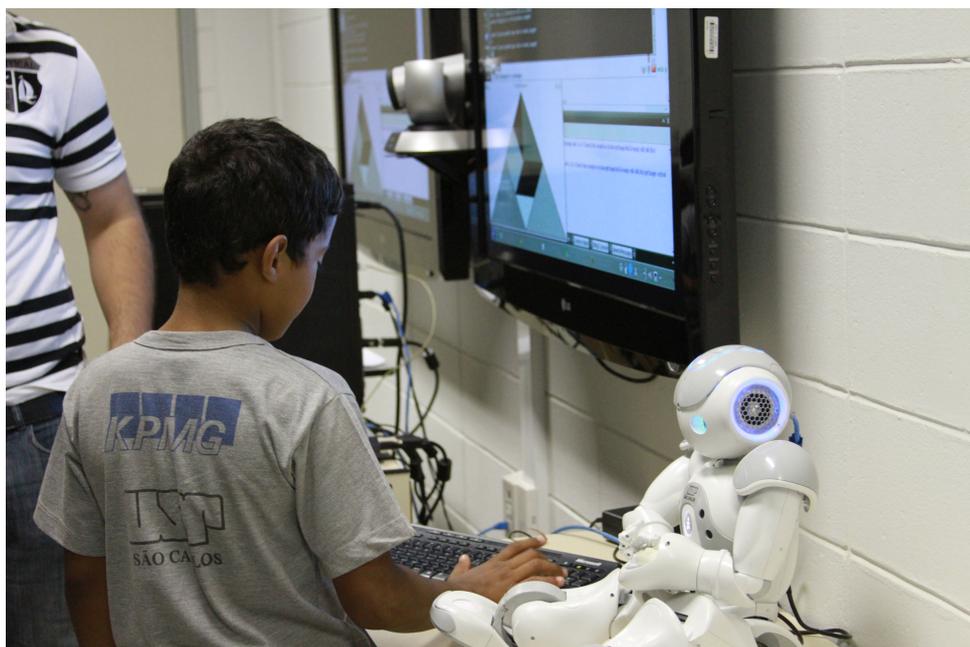


Figura 5.7: Aluno responde a uma pergunta do NAO

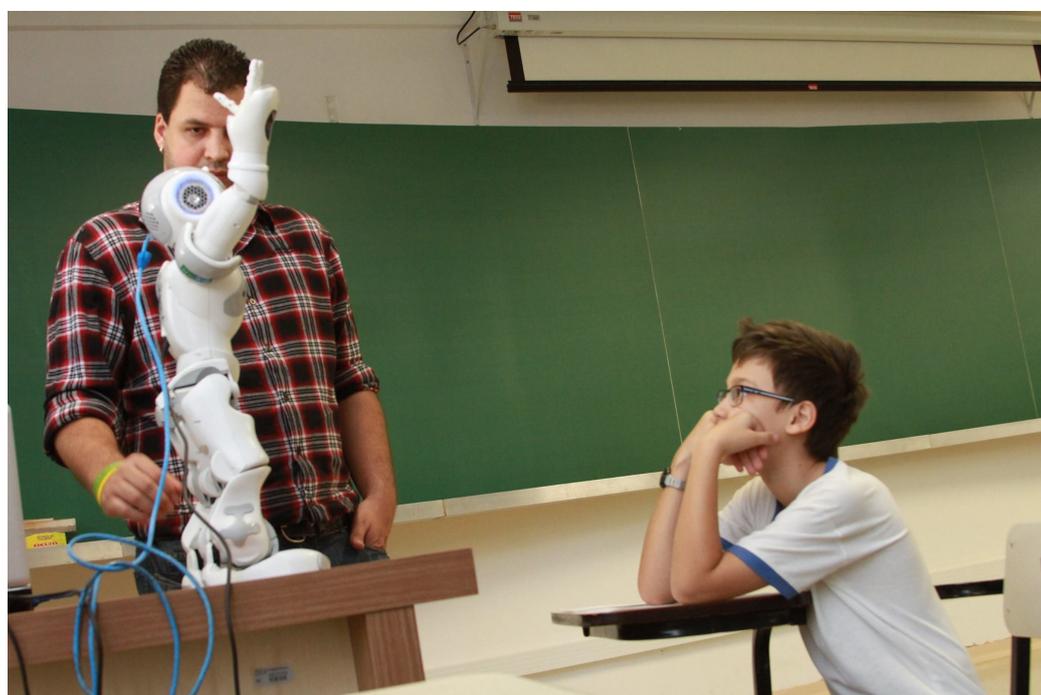


Figura 5.8: O NAO levanta os braços em sinal de alegria pela boa participação do aluno

- 2) Você conseguiu entender tudo o que o robô falou?
- 3) O robô deu alguma resposta errada durante a atividade?
- 4) Agora que você finalizou as atividades, o que você acha sobre a robótica?
- 5) Você acha que o robô pode ajudar com seus estudos?
- 6) O que você acha de ter um robô como professor?
- 7) O que você acha de tentarmos outras atividades com os robôs?

A Figura 5.9² mostra as crianças durante o ultimo estágio do experimento. Após responderem o questionário individualmente as crianças foram submetidas a um bate-papo em grupo, por meio do qual qualquer outra observação poderia ser levantada. Esse estágio foi realizado em função da timidez de alguns alunos em responder o questionário final.



Figura 5.9: Alunos em grupo discutindo as falhas e acertos do NAO durante todo o experimento

Os testes foram realizados em dias e horários diferentes, com 5 turmas de alunos diferentes. Foram realizados teste com quatro turmas de 15 alunos,

²Todas as crianças, professores e eventuais pais que participaram ou acompanharam os experimentos com o robô assinaram um termo de compromisso dando direito as imagens apresentadas nesta Dissertação e a todos os Artigos relacionados a esta pesquisa

uma turma de 12 alunos. Dessas turmas de alunos foi escolhido, aleatoriamente, os 10 alunos que formaram o grupo de controle, que executou as mesmas atividades aqui descritas, porém sem a utilização do robô NAO.

5.4 Discussão dos resultados

Em todos os experimentos foram feitas análises das respostas do robô por meio do autor desta Dissertação. Para medir a acurácia das respostas dadas, foi computada sempre a primeira resposta do robô para cada sujeito. Ou seja, quando o robô respondeu de forma errada (disse que a criança estava certa quando a resposta estava errada ou vice-versa) essa foi a resposta computada, mesmo que, após o reinício do experimento, ele tenha respondido de forma correta.

5.4.1 Atividade 1

Analisando os resultados do primeiro questionário, todos os participantes tinham o conhecimento prévio necessário para participar das atividades. O questionário tinha perguntas simples, e nenhum dos alunos declarou ter tido alguma dificuldade para completa-lo.

O mesmo aconteceu para o exercício do gato. Todas as crianças acertaram a quantidade de triângulos na primeira tentativa, sem precisar das dicas do robô NAO. E também todas as crianças encontraram as outras duas figuras geométricas na imagem, apesar de 70% delas responderem que era um losango (erro aceitável pela posição do quadrado na cena). É importante lembrar que o robô também não reconheceu as outras duas figuras, pois elas não faziam parte do escopo do reconhecimento para esta atividade. Como a primeira atividade com o robô era fácil, as crianças responderam positivamente a ela. Todas se mostravam felizes por terem acertado a pergunta do robô e, como eram os primeiros movimentos da atividade, prestavam total atenção ao que o robô fazia e dizia. A Tabela (5.1) mostra a Matrix de Confusão para este teste.

Como todas as crianças responderam o valor correto de triângulos, não obtivemos casos de Negativo Verdadeiro (quando a criança diz o valor errado e o robô confirma o erro dela) ou Falso Positivo (quando a criança diz o valor errado mas o robô reconhece como certo). Neste experimento, o robô teve um acerto de 80,65%, confirmando quando as crianças diziam que a imagem

Tabela 5.1: Matrix de Confusão - Teste 1

Matriz de Confusão	Verdadeiro	Falso
Positivo	50	0
Negativo	0	12

tinha 5 triângulos. Nos outros 19,35% das vezes, o robô não conseguiu reconhecer todos os triângulos, dizendo que a criança teria respondido errado. Isto demonstra que o sistema de reconhecimento proposto por [Benicasa \(2013\)](#) precisa ser aprimorado em termos de acurácia pois as constantes adotadas após os treinamentos da rede não foram 100% corretas.

5.4.2 Atividade 2

No segundo experimento, várias imagens de figuras geométricas eram espalhadas em frente ao aluno, enquanto o robô explicava como seria a atividade. Essas figuras espalhadas eram 3 quadrados, 3 triângulos, 2 paralelogramos, 2 retângulos, 1 trapézio, 1 elipse e 1 círculo. As figuras tinham cores sortidas (azul, amarelo, verde, vermelho) e tamanhos diferentes, para que a criança pudesse escolher. Nenhum padrão, entretanto, foi escolhido para as cores e tamanhos das figuras, que foram selecionadas a cada nova interação com o robô, garantindo-se apenas que teria, pelo menos, um triângulo, um quadrado e um retângulo.

Nesta atividade, apenas 4,8% dos estudantes conseguiram acertar todas as três figuras na primeira tentativa, o que fez o robô ter que dar dicas em todos os demais casos. O maior erro entre as crianças foi justamente a questão do quadrado, pois uma vez que conseguiam acertar o retângulo na primeira pergunta, não prestavam atenção na última, e mostravam a mesma figura. Esse cenário foi intencional, justamente para apresentar para as crianças a importância de se ouvir direto a pergunta. Neste momento, algumas vezes foi necessário intervir no teste, explicando o que tinha acontecido e reforçando o pedido para se ouvisse atentamente as palavras do NAO. Como o robô não estava preparado para repetir a pergunta inicial após uma resposta errada, o mediador do experimento teve que explicar o que o robô tinha falado. Esse problema aconteceu em 19.35% dos casos.

No final do teste, alguns alunos se queixaram de não conseguir ouvir o que o NAO dizia neste experimento, e disseram que ficaram confusos com a escolha da figura certa. Isso foi comprovado pela escolha, em alguns casos,

de mostrar o círculo ou elipse, por ouvirem erroneamente o robô dizer "figura com **um** lado" ao invés de "figura com **quatro** lados". A descrição do MOS irá demonstrar o grau de satisfação dos alunos com as respostas do robô. Com relação a atenção visual, o robô conseguiu, em todos os testes uma acurácia de 79% no reconhecimento correto das figuras mostradas pelos alunos. Nos casos em que ocorriam falsos positivos ou negativos, o experimento era interrompido, o erro do robô era explicado para o aluno, e o experimento era retomado. A Tabela 5.2 mostra a Matriz de Confusão correspondente a este experimento.

Tabela 5.2: Matrix de Confusão - Teste 2

Matriz de Confusão	Verdadeiro	Falso
Positivo	109	55
Negativo	242	49

Este foi o experimento no qual o modelo proposto foi mais vezes testado. Contando todas as dicas dadas, o robô fez um total de 445 perguntas durante todo o experimento com os 62 alunos. O alto número de Verdadeiros Negativos mostra a quantidade de vezes que as crianças erraram a resposta esperada e o robô percebeu este erro. Ainda teve uma quantidade bastante considerável de Falso Positivo, quando a criança errou a imagem, mas o robô reconheceu como certa. A maior parte desses erros aconteceu justamente no reconhecimento do quadrado e do retângulo, dada a similaridade das figuras. 58,18% dos Falso Positivos ocorreram nesses casos.

5.4.3 Atividade 3

O terceiro experimento foi um jogo que media a capacidade das crianças e do robô em reconhecer figuras geométricas sobrepostas. A ideia principal do experimento foi propiciar que a criança pudesse perceber que duas figuras juntas podem formar uma outra figura, que estaria "escondida" na imagem, porém essa informação não era passada para as crianças a princípio. O robô deliberadamente tinha um tom mais desafiador neste experimento, instigando as crianças a participar. Do caso da primeira imagem (Figura 5.5(a)) nenhuma das crianças acertou a quantidade de figuras que o robô encontrou (existiam 14 possíveis). Em todos os casos, a primeira resposta das crianças foi 7 figuras, que correspondiam as peças e se mostravam muito surpresas quando o robô dizia que tinha encontrado mais figuras.

Sem uma dica mais concreta (pois nem o robô, nem o mediador davam dicas para esta atividade) os alunos se empenharam muito para tentar encontrar as figuras escondidas, mas 83,87% deles falharam em conseguir encontrá-las, e os outros 16,12% até encontraram mais alguma, mas nenhuma delas conseguiu encontrar mais do que 9 figuras. Após responderem a segunda vez de forma errada, o robô mostrava as figuras escondidas, mostrando-as de forma destacada (como na Figura 5.5(b)).

Após a explicação do robô, ele pedia a segunda imagem (Figura 5.6), e novamente pedia a quantidade de imagens escondidas. Ficou claro para os estudantes, neste segundo momento, como funcionava o jogo e todos relutaram em dar uma resposta muito rápida ou voltar a responder que o número de figuras encontradas era 7. Empenhados em vencer o robô, todos encontraram pelo menos uma figura escondida já na primeira tentativa. 8,06% deles conseguiram encontrar a mesma quantidade de figuras que o robô encontrou na segunda tentativa. O robô, como dito no Capítulo 4, não estava preparado para reconhecer todas as figuras no início (a rede MLP estava limitada, então não rotulava de forma correta), o que fez com 4,83% dos alunos conseguissem, inclusive, vencer o robô, encontrando figuras que o NAO não tinha encontrado.

A percepção das crianças sobre as figuras geométricas aumentou durante os experimentos realizados. Isso fica claro ao compararmos as respostas obtidas no início e no final do teste, com o segundo questionário. Enquanto no começo da terceira atividade nenhum aluno encontrou as figuras sobrepostas, no questionário todas as crianças conseguiram encontrar o valor esperado de figuras, e alguns encontraram até mais possibilidades do que foi previsto. Na conversa após as atividades, crianças e professores estavam satisfeitos com a atividade conduzida pelo robô NAO. A Tabela 5.3 apresenta a Matrix de Confusão da ultima atividade.

Tabela 5.3: Matrix de Confusão - Teste 3

Matriz de confusão	Verdadeiro	Falso
Positivo	0	5
Negativo	240	3

Esta foi a atividade que mais atraiu as crianças, pois era a mais complicada e a mais desafiante, porém foi a mais complicada para que o robô reconhecesse corretamente. Isso mostrou que o sistema proposto por

Benicasa (2013) precisa ser aprimorado para este tipo de tarefa. Não houve Verdadeiros positivos, pois apenas três crianças acertaram a quantidade correta de figuras, e em todas o robô tinha encontrado um valor menor. Em 5 casos, as crianças encontraram a mesma quantidade de imagens reconhecidas pelo robô, configurando os Falsos Positivos. Ainda assim, o robô teve uma boa medida de Verdadeiros Negativos, percebendo quando as crianças encontraram menos figuras do que o esperado.

5.4.4 Avaliação dos Alunos

No final dos testes realizados, os alunos tiveram espaço para expor suas opiniões sobre a atividade conduzida pelo robô NAO. O resultado da avaliação MOS pode ser visto na Tabela 5.4.

Tabela 5.4: Tabela de Avaliação MOS

MOS (%)	Excelente	Bom	Regular	Ruim	Péssimo
1	40.3	25.84	24.19	0	9.67
2	79	21	0	0	0
3	64.5	19.35	1.61	3.22	11.29
4	88.7	4.83	3.22	3.22	0
5	50	16.12	9.67	24.21	0
6	6.48	25.8	32.25	20.96	14.51
7	100	0	0	0	0

Apesar de 100% dos alunos terem dito que participariam de mais experimentos com os robôs, 34% deles tem bastantes dúvidas quando a possibilidade de ter um robô como professor. Além da necessidade do mediador para garantir que o NAO funcione, alguns outros problemas apontados pelos próprios alunos é a falta de investimentos na robótica: 59.67% apontaram que existe muito mais escolas do que robôs e 95.16% disseram que o valor dos robôs ainda está muito alto para ser utilizado nas escolas públicas. As falhas do robô (erros de reconhecimento, problemas com a fala, dificuldades de movimento [uma vez que o robô quase caiu durante alguns testes]) foram apontados por 77.41% dos alunos como um grande impedimento para o uso do robô na sala de aula.

Durante o experimento, foi observado que os estudantes se sentiam motivados para realizar as tarefas solicitadas pelo robô. E em tom de brincadeira, os alunos acabaram absorvendo o conteúdo que estava sendo

exposto, sem nem ao menos considerar toda a experiência como aula.

Isso fica muito claro ao comparar os alunos que fizeram os experimento com o robô e os que fizeram as atividades com um professor humano. Os alunos taxaram as atividades de chatas, e logo perdiam o interesse pelo que estava sendo ministrado, principalmente pelo fato de ser um professor humano e diferente daquele que eles estavam acostumados em suas escolas (já que o professor acompanhava as atividades com o NAO).

A partir da segunda atividade os alunos interromperam o professor durante a explicação e, na terceira atividade (figuras sobrepostas) os alunos se empenharam bem menos para responder a quantidade de figuras encontradas. Enquanto a média de tempo da terceira atividade foi de 6,3 minutos para dar uma resposta, o grupo de controle normalmente tentava adivinhar a questão em 2,6 minutos. É notório que a motivação, no contexto da aprendizagem, é fator importante para que o aluno consiga absorver o conteúdo.

5.5 Considerações Finais

Neste capítulo, foram apresentados os vários experimentos realizados com crianças, visando testar o sistema integrado desenvolvido e incorporado ao robô NAO. A partir de alguns jogos de perguntas e respostas, o robô NAO desafiou as crianças que participavam do teste, competindo com elas, graças ao sistema implementado nesta dissertação. Analisando o algoritmo proposto, foi obtida uma acurácia de 79.09%. Já em relação ao robô em sala de aula, as atividades foram muito bem recebidas por professores e alunos, que gostaram da possibilidade de se ter um robô como este inserido em sala de aulas. Além da disciplina e dedicação durante os testes com os robôs, os alunos mostraram nos questionários preenchidos que o conteúdo da aula tinha sido absorvido.

Conclusão e Trabalhos Futuros

6.1 Conclusão

A tecnologia têm avançado rapidamente, tornando-se mais abrangente a cada dia. Esse crescimento de possibilidades nos permite propor técnicas alternativas para o ensino. O objetivo deste trabalho foi fornecer a um robô humanoide a capacidade de reconhecer objetos, utilizando um sistema de atenção visual proposto por [Benicasa \(2013\)](#), e com isso interagir com crianças, em algumas atividades visando o reforço e a fixação dos estudos sobre figuras geométricas. O sistema de atenção visual é constituído por quatro fases principais: a segmentação da imagem através de rede LEGION, o reconhecimento do objeto utilizando uma rede MLP e o uso da saída da MLP e de informações primitivas da imagem em um mapa SOM para se decidir o ponto de maior saliência.

Com o uso de inibidores globais e osciladores locais, a rede LEGION separa regiões homogêneas, que representam objetos da imagem, das informações heterogêneas, que representam o fundo da cena ou outras informações irrelevantes para o escopo deste trabalho. Apesar da implementação utilizada ter apenas dois parâmetros para ajuste, foram necessários muitos testes para se encontrar os valores que melhor segmentavam a imagem recebida pela câmera do robô.

Para se treinar a rede MLP, foi necessário um bom conjunto de treinamento com imagens já segmentadas pela LEGION. Uma vez que não existe um padrão

de como as crianças vão mostrar as imagens para o robô, podendo ser mais perto ou mais longe da câmera, este conjunto de teste precisou ser muito bem definido para que a rede pudesse reconhecer os objetos. Uma vez que a rede reconhecia o objeto segmentado, essa informação era usada para guiar a atenção.

O mapa SOM foi treinado para gerar o mapa de saliência e definir o ponto de maior interesse na cena, com base nos dois processos anteriores. Com esta informação, o robô ignora o resto da cena, e mostra para o aluno apenas o objeto considerado o de maior atenção. Graças a inibidores de retorno, uma vez reconhecido um objeto, o robô volta sua atenção para o próximo ponto de atenção e assim sucessivamente até se esgotar os segmentos encontrados e reconhecidos.

Esse sistema foi adaptado para a utilização no robô NAO, considerando-se características de hardware (como a adaptação para sua câmera) e características de software. Foi necessário utilizar uma versão específica da biblioteca OpenCV para que ela fosse compatível com o modelo do robô presente no LAR/USP. Toda a interação do sistema com as crianças era feita através do robô, então alguns comportamentos foram criados para as diversas possíveis situações durante os experimentos, como demonstrar alegria ou tristeza de acordo com o desempenho do sujeito, como uma forma de humanização do humanoide utilizado.

A avaliação deste sistema proposto foi feita em sala de aula, com alunos de diversas escolas da cidade de São Carlos. Três atividades com o uso do robô foram propostas, com diferentes níveis de dificuldade, com a ajuda de profissionais da área de educação matemática. Estas atividades incluíam o reconhecimento de figuras geométricas, conhecimento de algumas propriedades como fórmulas de área e perímetro das figuras e a compreensão de figuras sobrepostas. Todos os experimentos foram preparados para alunos do terceiro e quarto ano do ensino fundamental II.

O sistema foi avaliado por meio da acurácia obtida, verificando-se a quantidade de vezes que a resposta do robô foi coerente com a informação passada pelo aluno durante o experimento, e também foi avaliado subjetivamente (através da técnica de avaliação MOS) pelos alunos e professores que participaram da atividade. Com relação a acurácia o sistema foi satisfatório, respondendo de maneira correta na maioria dos casos (79.09%), corroborando os resultados obtidos e apresentados em [Benicasa \(2013\)](#). Com relação a avaliação subjetiva, todos os alunos disseram que

participariam novamente de trabalhos utilizando robôs, e se mostraram muito satisfeitos com os trabalhos propostos. Apesar de apontarem algumas falhas, como os problemas de entender a fala do robô e até os momentos em que o robô errou no reconhecimento, os alunos que participaram da atividade com o robô tiveram um desempenho melhor no questionário final (84.61% de acertos) do que aqueles que fizeram as atividades sem o robô (60% de acertos).

Os experimentos apresentaram resultados promissores, mas são necessários mais testes para se obter uma conclusão definitiva do uso deste robô em sala de aula. De qualquer modo, o sistema desenvolvido nesta dissertação de Mestrado, que capacita o robô para reconhecer figuras geométricas planares e interagir com crianças sobre este conteúdo, é inovador, pois, pelo que é do nosso conhecimento, não existe ainda tal sistema. Este trabalho colabora com as demais pesquisas de robótica educacional do grupo de pesquisa do LAR/USP, que darão continuidade a esta pesquisa.

Os seguintes artigos foram gerados, publicados e apresentados durante esta dissertação de Mestrado:

- PINTO, A. H. M. ; Benicasa, A.X. ; OLIVEIRA, L. O. ; MENEGHETTI, R. C. G. ; Romero, Roseli A.F. . Attention Based Object Recognition applied to a Humanoid Robot. In: LARS/SBR 2014, 2014, São Carlos. Proceedings of LARS/SBR 2014, 2014.
- PINTO, A. H. M. ; ORLANDI, L. ; MENEGHETTI, R. C. G. ; ROMERO, R. A. F. . Inserção de um robô humanoide no Ensino de Objetos Geométricos 2D sobrepostos. In: 25o. SBIE Trilha 5, 2014, Dourados - MS. Anais do 25o. SBIE, 2014.
- PINTO, A. H. M. ; OLIVEIRA, L. O. ; BENICASA, ALCIDES X. ; MENEGHETTI, R. C. G. ; ROMERO, R. A. F. . Incorporating a Humanoid Robot to Motivate the Geometric Figures Learning. In: TISE2014 - XIX Conferência Internacional sobre Informática na Educação, 2014, Fortaleza - CE. Proceedings, 2014.

6.2 Trabalhos Futuros

Como continuidade deste trabalho, pretende-se melhorar os principais problemas percebidos durante o teste e também apontados pelas crianças. É

necessário aprimorar os ajustes dos parâmetros das redes LEGION e MLP, de forma a obter uma acurácia ainda melhor no reconhecimento de figuras geométricas 2D. Uma possibilidade para este ajuste, pode ser usar algoritmos genéticos.

Uma outra extensão possível é aperfeiçoar o sistema desenvolvido para que possa reconhecer figuras geométricas 3D.

Além disso, é necessário aprimorar o sistema para o reconhecimento de figuras sobrepostas. Como visto durante os experimentos, o sistema se equivocou no reconhecimento em alguns testes, o que pode ter sido causado por um conjunto de treinamento insuficiente ou pela dificuldade de segmentação no caso apresentado.

Com a proposta de se propiciar uma interação mais natural possível, é necessário ajustar o reconhecimento de voz, para que o aluno não precise utilizar o computador para se comunicar com o robô. Isso irá incrementar a experiência dos alunos, que podem se sentir ainda mais motivados com as atividades e mais a vontade para interagir com o robô.

Para ampliar o número de atividades para os alunos, conseguir que o robô reconheça as peças do Tangram, por exemplo, pode ampliar muito as possibilidades de aplicações e atividades com os alunos.

Referências Bibliográficas

- Ashby, W. R. (1952). *Design for a Brain: The Origin of Adaptive Behaviour*. Wiley. Citado na página 5.
- Ashby, W. R. (1956). *An Introduction to Cybernetics*. Chapman and Hall. Citado na página 5.
- Beauchamp, C. (1998). Revolution industrielle et croissance économique au XIX^e siècle. página 70. Citado na página 1.
- Benicasa, A., Quiles, M., Zhao, L., e Romero, R. (2013). Top-down biasing and modulation for object-based visual attention. In Lee, M., Hirose, A., Hou, Z.-G., e Kil, R., editors, *Neural Information Processing*, volume 8228 of *Lecture Notes in Computer Science*, páginas 325–332. Springer Berlin Heidelberg. Citado nas páginas 15, 16, e 43.
- Benicasa, A. X. (2013). *Sistemas Computacionais para Atenção Visual Top-Down e Bottom-Up usando Redes Neurais Artificiais*. Tese de Doutorado, Universidade de São Paulo - São Carlos. Citado nas páginas xi, 16, 25, 26, 27, 28, 35, 37, 38, 42, 44, 46, 60, 63, 65, e 66.
- Bezdek, J. C., Pal, D., e K., S. (1992). *Fuzzy models for pattern recognition: methods that search for structures in data*. New York: IEEE Press. Citado na página 9.
- Bjoorne, P. e Balkenius, C. (2005). A model of attentional impairments in autism: First steps toward a computational theory. *cognitive systems research*, 6(3):193-204. Citado na página 5.
- Blank, D. S. e Ross, J. O. (1997). Incorporating a connectionist vision module into a fuzzy behavior-based robot controller. In *Proceedings of the 1997*

Midwest Artificial Intelligence and Cognitive Science Society Cybernetcs. Citado na página 11.

Borji, A., Ahmadabadi, M. N., e Araabi, B. N. (2011). Cost-sensitive learning of top-down modulation for attentional control. *Machine Vision and Applications* 22(1), 61-76. Citado na página 14.

Borji, A. e Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(1), 185-207. Citado nas páginas 13 e 14.

Breazeal, C. (2002). *Designing sociable robots*. The MIT Press. Citado na página 5.

Brooks, R. A. (1985). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, Vol.2, No.1, pg 14 a 23, MIT AI Memo 864. Citado na página 1.

Burt, P. J. e Adelson, E. H. (1983). The laplacian pyramid as a compact image code. Citado na página 26.

Campbell, S. R., Wang, D. L., e Jayaprakash, C. (1999). Synchrony and desynchrony in integrate-and-fire oscillators. *Neural Computation* 11, 1595-1619. Citado na página 16.

Chakraborty, A. K., Pal, D., e Chatterjee, P. (2012). Fast recognition of mechanical objects using neural networks under robust aspect. *Journal of The Institution of Engineers (India)* vol. 93(1), pp. 55-62. Citado nas páginas 9 e 12.

Chatterjee, P., Karan, B. M., e Sinha, P. K. (2002). Multi-layer feed-forward modular network for induction motor. *IEEE World Congress on Computational Intelligence*, vol. 1, pp.339. Citado na página 12.

Chavez, G. C. (2002). *Sistema celular evolutivo para reconhecimento de padrao invariante*. Universidade de Sao Paulo. Citado nas páginas xi e 3.

Cheng, H. D., Jiang, X., Sun, Y., e Wang, J. (2001). Color image segmentation: advances and prospects. *pattern recognition* (34), 2259-2281. Citado nas páginas 20, 22, 23, e 24.

- Choksuriwong, A., Laurent, H., e Emile, B. (2005). Comparison of invariant descriptors for object recognition. In: IEE International Conference On Image Processing. pp. 377-380. Citado na página 7.
- Conci, A., Azevedo, E., e Leta, F. R. (2008). Computacao grafica: Teoria e pratica. v2, rio de janeiro: Campus. Citado nas páginas 7 e 8.
- Crabtree, J. S., Ehrlich, R., e Prince, C. (1984). Evaluation of strategies for segmentation of blue-dyed pores in thin sections of reservoir rocks. *Computer Vision, Graphics, and Image Processing*, 28:1-18. Citado nas páginas xi e 23.
- Csala, E., Nemeth, G., e Zainko, C. (2012). Application of the nao humanoid robot in the treatment of marrow-transplanted children. *International Conference on Cognitive Infocommunications (CogInfoCom)*, páginas pp. 655-659. Citado na página 16.
- de Padua Braga, A., Ludermir, T. B., e de Leon Ferreira Carvalho, A. C. P. (2000). *Redes neurais artificiais: Teoria e aplicacoes*. Itc. Citado nas páginas 29 e 33.
- Desimone, R. e Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience* 18, 193-222. Citado na página 14.
- Duffy, B. R., Rooney, C., O'Hare, G., e O'Donoghue, R. (1999). What is a social robot? in 10th irish conf. on artificial intelligence and cognitive science, ucc, ireland. Citado na página 5.
- Egmont-Petersen, M., de Ridder, D., e Handels, H. (2002). Image processing with neural networks - a review. *Pattern Recognition* (35), 2279-2301. Citado na página 9.
- Elazary, L. e Itti, L. (2010). A bayesian model for efficient visual search and recognition. *Vision Research* 50(14), 1338-1352. Citado na página 14.
- ENVI (2000). *Guia do envi em portugues*. Sulsoft, 2000. Diponivel em www.sulsoft.com.br. Citado nas páginas xi e 22.
- Fiaz, M. K. e Ijaz, B. (2010). Vision based human activity tracking using artificial neural networks. *Internacional Conference on Intelligent and Advanced Systems (ICIAS)*, páginas pp. 1 - 5. Citado na página 12.

- Flusser, J. e Suk, T. (1996). Pattern recognition by affine moment invariants. *Pattern Recognition*. vol 26: pp. 167-174. Citado na página 8.
- Gaspar, T. L. (2006). Reconhecimento de faces humanas usando redes neurais mlp. Dissertacao, Escola de Engenharia de Sao Carlos, USP. Citado na página 2.
- Gonzalez, R. C. e Woods, R. E. (2002). *Procesamento de imagens digitais*. Sao Paulo-SP: Edgard Blucher LTDA. Citado nas páginas 7 e 23.
- Harel, J., Koch, C., e Perona, P. (2006). Graph-based visual saliency. In *Advances in neural information processing systems*, pp. 545-552. Citado na página 14.
- Haykin, S. (2001). *Artificial Neural Networks: Methods and Application*. Bookman. Citado nas páginas 2, 31, e 37.
- Itti, L. e Koch, C. (2001). Computational modelling of visual attention. Citado nas páginas 4 e 13.
- Itti, L., Koch, C., e Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(11), 1254-1259. Citado nas páginas 14 e 15.
- Iyengar, S. e Kashyap, R. (1991). Neural networks: a computational perspective. In ANTOGNETTI, Paolo, MILUTINOVIC, Veljko. *Neural networks: concepts, applications, and implementations*. Vol II, pp. 1-30. Englewood Cliffs: Prentice Hall. Citado na página 9.
- Jahne, B. (1997). *Practical handbook on image processing for scientific application*. CRC Press. Citado na página 22.
- Jain, A., Duin, R., e Mao, J. (2000). Statistical pattern recognition: A review. *IEEE transactions on pattern analysis and machine intelligence*, v.22, n.1, p. 4-37. Citado na página 2.
- Kandel, E. R., Schwartz, J. H., e Jessell, T. M. (1997). *Fundamentos da neurociencia e do comportamento*. guanabara koogan. Citado nas páginas 29 e 30.
- Kazanovich, Y. B. e Borisyuk, R. (2002). Object selection by an oscillatory neural network. *BioSyst.* 67, 103-111. Citado na página 16.

- Khatchatourian, O. e Padilha, F. R. R. (2008). Reconhecimento de variedades de soja por meio do processamento de imagens digitais usando redes neurais artificiais. *eng. agric., jaboticabal*, v. 28, n. 4, dez. Citado na página 2.
- Kimberlee, J., King, M., Hellersteth, S., Wiren, A., e Mulligan, H. (2013). Feasibility of using a humanoid robot enhancing attention and social skills in adolescents with autism spectrum disorder. *International Journal of Rehabilitation Research*. Citado na página 16.
- Koch, C. e Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology* 4, 219-227. Citado nas páginas 13 e 27.
- Koenemann, J. e Bennewitz, M. (2012). Whole-body imitation of human motions with a nao humanoid. presented at the Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, Boston, Massachusetts, USA. Citado na página 17.
- Kohonen, T. (2001). *Self-organizing maps* (3th edition ed.). Citado nas páginas 25, 37, e 38.
- Kurweil, R. (1990). *The age of spiritual machines*. Dissertação de Mestrado, MIT Press. Citado na página 1.
- LindBlad, T. K. J. (2005). *Image processing using pulse-coupled neural network*. Secaucus, NJ, USA: Springer. Citado na página 11.
- Liu, G. e Yang, J. (2008). Image retrieval based on the texton co-occurrence matrix. *Pattern Recognition*. Citado na página 8.
- Long, F., Zhang, H., e Feng, D. (2005). Fundamentals of content-based image retrieval. *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications (Signals and Communication Technology)*, pp.1-26. Citado na página 8.
- Lucchese, L. e Mitra, S. K. (2001). Color image segmentation: A state-of-the-art-survey. In *Proc. of the Indian National Science Academy (INSA-A)*. Volume 67 of *Image Processing, Vision and Pattern Recognition*, New Delhi, India. pp.207-221. Citado na página 24.
- Luger, G. F. (2004). *Artificial Intelligence - Structures and Strategies for Complex Problem Solving*. Pearson Education. Citado na página 1.

- Machado, A. (2000). Neuroanatomia funcional (2.ed. ed.). rio de janeiro: Atheneu. Citado na página 29.
- Moik, J. G. (1980). Digital processing of remotely sense images. u.s. government printing office, washington dc, eua. Citado na página 2.
- Nalwa, V. S. (1993). A guided tour for computer vision. Addison-Wesley. Citado na página 22.
- Ogawa, T. e Komatsu, H. (2004). Target selection in area v4 during a multidimensional visual search task. Journal of Neuroscience 24(28), 6371-6382. Citado na página 13.
- Ong, S. H., Yeo, N. C., Lee, K. H., Venkatesh, Y. V., e Cao, D. M. (2002). Segmentation of color images using a two-stage self-organizing network. Image and Vision Computing (20), 63-68. Citado na página 9.
- Papamarkos, N. (1999). Color reduction using local features and a sofml neural network. Int. Journal of Imaging Systems and Technology 10, pp. 404-409. Citado na página 10.
- Papamarkos, N., Atsalakis, A., e Strouthpoulos, C. (2002). Adaptive color reduction. IEEE Trans. on Systems, Man and Cybernetics-Part B 32(1), pp.44-56. Citado na página 10.
- Papert, S. (1994). *The Childrens's Machine: Rethinking School in the age of the computer*. Perseus Books. Citado na página 5.
- Pedrin, H. e Schwartz, W. R. (2008). Análise de imagens digitais: Princípios, algoritmos e aplicações. São Paulo: Ed. Thomson. Citado na página 8.
- Pomerleau, D. (1995). Neural network vision for robot driving. In M. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks*. Citado na página 11.
- Quiles, M. G. (2004). Sistema de visão baseado em redes neurais para o controle de robôs móveis. Dissertação de Mestrado. Citado nas páginas xi, 11, 19, 21, 31, 32, e 34.
- Quiles, M. G., Romero, R. A. F., e Zhao, L. (2006). A pulse-coupled neural network as a simplified bottom-up visual attention model. In IEEE Proceedings of the Ninth Brazilian Symposium on Artificial Neural Networks (SBRN 2006), pp. 1-6. Citado na página 11.

- Quiles, M. G., Wang, D., Zhao, L., Romero, R. A., e Huang, D. (2011). Selecting salient objects in real scenes: An oscillatory correlation model. *Neural Networks* 24(1), 54-64. Citado na página 16.
- Rec, I. (2000). *P.800: Methods for subjective determination of transmission quality. CCITT Recommendations; electronic version*. Recommendations. IUT, Geneva. Citado na página 50.
- Robins, B., Dickerson, P., Stribling, P., e Dautenhahn (2004). Robot-mediated joint attention in children with autism: A case study in robot-human interaction. *interaction studies*, 5(2):161-198. Citado na página 5.
- Robotics, A. (2014). Nao software 1.14.5 documentation. Citado nas páginas xi e 42.
- Russel, S. e Norvig, P. (2004). *Inteligencia Artificial: Uma abordagem moderna*. Ed. Campus, Sao Paulo. Citado na página 2.
- Rutishauser, U., Walther, D., Cock, C., e Perona, P. (2004). Is bottom-up attention useful for object recognition? In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, Volume 2*, pp. II-37. Citado na página 14.
- Sakar, P. (2000). A brief history of cellular automata. *ACM Computing Surveys*, v. 32, n. 1, p.80-107. Citado na página 8.
- Shamsuddin, S., Yussof, H., Ismail, L., Hanapiah, F. A., Mohamed, S., Piah, H. A., e Zahari, N. I. (2012). Initial response of autistic children in human-robot interaction therapy with humanoid robot nao. *International Colloquium on Signal Processing and its Applications*, páginas pp. 188–193. Citado na página 17.
- Shic, F. e Scassellati, B. (2007). A behavioral analysis of computational models of visual attention. Citado na página 4.
- Simoese, A. S. e Costa, A. H. R. (2000a). Segmentacao de imagens por classificacao de cores: Uma abordagem neural para representacao rgb. In C. H. C. R. M. T. S. Sakude (Ed.), *Workshop de computacao WORKCOMP2002*, ITA-Sao Jose dos Campos, pp. 25-31. Citado na página 10.
- Simoese, A. S. e Costa, A. H. R. (2000b). Using neural color classification in robotic soccer domain. In *Internacional Joint Conference IBERAMIA2000*. Citado na página 10.

- Simoese, A. S. e Costa, A. H. R. (2001). Classificacao de cores por redes neurais artificiais: Um estudo do uso de diferentes sistemas de representacao de cores no futebol de robos moveis autonomos. In ENIA. Citado na página 10.
- Skarbek, W. e Koschan, A. (1994). Colour image segmentation - a survey. Technical Report, Universidade de Berlin. Citado na página 23.
- Smolar, P., Tuharsky, J., Fedor, Z., Vircikova, M., e Sincak, P. (2011). Development of cognitive capabilities for robot nao in center for intelligent technologies in kosice. In *Cognitive Infocommunications (CogInfoCom), 2011 2nd International Conference on*, páginas 1–5. IEEE. Citado na página 18.
- Spasojevic, S. S., Susic, M. Z., e Durovic, Z. M. (2012). Recognition and classification of geometric shapes using neural networks. *11th Symposium on Neural Network Application in Electrical Engineering, NEUREL-2012*, páginas pp. 71–76. Citado na página 12.
- Starna, A. (2009). *Compositing digitale*. Cinema, Media. Citado nas páginas xi e 20.
- Suay, H. B. e Chernova, S. (2011). Humanoid robot control using depth camera. presented at the Proceedings of the 6th international conference on Human-robot interaction, Lausanne, Switzerland. Citado na página 17.
- Tanaka, F. e Ghosh, M. (2011). The implementation of care-receiving robot at an english learning school for children. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*, páginas 265–266. IEEE. Citado na página 18.
- Tapus, A., Peca, A., Aly, A., Pop, C., Jisa, L., Pintea, S., Rusu, A. S., e David, D. O. (2012). Childrens with autism social engagement in interaction with nao, an imitative robot - a series of single case experiments. *John Benjamins Publishing Company*, vol. 13(3):pp. 315–347. Citado na página 16.
- Tejada, C. Z. F. (2008). Analisis de metodos de procesamiento de imagenes estereoscopicas forestales. mestrado em investigacao em informatica. departamento de ingenieria del software e inteligencia artificial - facultad de informatica - universidad complutense - madrid, espanha. Citado na página 2.

- Tsai, S. e Tseng, Y. (2012). A novel color detection method based on hsl color space for robotic soccer competition. *Computers and Mathematics with Applications*, 64:1291–1300. Citado na página 10.
- Tsotos, J. K. (2011). A computational perspective on visual attention. MIT Press, Cambridge. Citado na página 13.
- Tzafestas, C., Mitsou, N., Georgakarakos, N., Diamanti, O., Maragos, P., Fotinea, S. E., e Efthimiou, E. (2009). Gestural teleoperation of a mobile robot based on visual recognition of sign language static handshapes. *IEEE International Symposium on Robot and Human Interactive Communication*, páginas pp. 1073–1079. Citado na página 11.
- Veltkamp, R. C. e Latecki, L. J. (2006). Properties and performance of shape similarity measures. In: *International Conference on Data Science and Classification*, pp. 1-9. Citado na página 8.
- Veltrop, T. (2012). Disponível em: <http://taylor.veltrop.com/>. Citado na página 17.
- Verner, I. M., Polishuk, A., Klein, Y., Cuperman, D., Mir, R., e Wertheim, I. (2012). A learning excellence program in a science museum as a pathway into robotics. *International Journal of Engineering Education*, 28(3):523. Citado na página 17.
- von der Malsburg, C. (1981). The correlation theory of brain function. Technical report, Internal report 81-2: Max-Planck Institute for Biophysical Chemistry, Göttingen, Germany. Citado na página 15.
- Waldherr, S., Thrun, S., e Romero, R. A. F. (2000). A gesture-based interface for human-robot interaction. *Autonomous Robots* 9(2), 151-173. Citado na página 11.
- Walther, D., Itti, L., Riesenhuber, M., Poggio, T., e Koch, C. (2002). Attentional selection for object recognition-a gentle way. Citado na página 14.
- Walther, D., Rutishauser, U., Cock, C., e Perona, P. (2005). Selective visual attention enables learning and recognition of multiples objects in cluttered scenes. *Computer Vision and Image Understanding* 100, 41-63. Citado na página 14.
- Wang, D. (2005). The time dimension for scene analysis. *IEEE Transactions on Neural Networks* 16(6), 1401-1426. Citado na página 15.

- Wang, S. (1995). The unpredictability of standard back propagation neural networks in classification applications. *Management Science*, vol. 41: pp. 555-559. Citado nas páginas [xi](#), [9](#), [15](#), [34](#), [35](#), [36](#), e [44](#).
- Wolfe, J. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin Review* 1, 202-238. Citado nas páginas [xi](#), [14](#), e [15](#).
- Zaman, B., Jatmiko, W., Wibowo, A., e Imah, E. M. (2011). Implementation vehicle classification on distributed traffic light control system neural network based. *Internacional Conference Advanced Computer Science and Information System (ICACSIS)*, páginas pp. 107–112. Citado na página [12](#).
- Zhang, D. e Lu, G. (2004). Review of shape representation and description techniques. *Patter Recognition*, v.37, n.1, p. 1-19. Citado na página [7](#).
- Zitova, B. e Flusser, J. (2003). Image registration methods: A survey. *Image and Vision Computing*, vol 21, pp.977-1000. Citado na página [8](#).